

ABSTRACT

Title of dissertation: WIRELESS SENSING FOR
VITAL SIGNS MONITORING

Fengyu Wang
Doctor of Philosophy, 2021

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

Continuous monitoring of vital signs (e.g., respiration and heart rate, heart rate variability, etc.) is critical for early detection and prevention of potentially fatal diseases. Existing solutions usually require users to wear dedicated devices such as wrist-worn sensors or chest straps to physically contact with human body, which is uncomfortable for users and sometimes may cause skin allergies. With the rapid development of the Internet of Things (IoT), wireless sensing has gained increasing attention in recent years because of the ubiquitous deployment of wireless devices. It has been proved that the presence of human will affect wireless signal propagation, enabling the functionality of wirelessly monitoring human subjects by analyzing the electromagnetic (EM) wave.

Despite of the wide variety of IoT devices, most of them are equipped with WiFi, which is a very mature and cost-effective connectivity solution. Moreover, as the next-generation wireless communication technique, millimeter wave (mmWave) radio has become available on home routers, vehicles, etc., to achieve

higher performance (e.g., larger bandwidth, higher directionality). Motivated by the increasing demand of monitoring vital signs as well as the development of IoT, in this dissertation, we propose four wireless sensing systems to monitor vital signs leveraging the channel information of commercial devices.

In the first part, using the Channel State Information (CSI) of a single pair of commercial WiFi devices, a novel system is proposed to continuously track the breathing rates of multiple persons without requiring prior knowledge of crowd number. By leveraging both the spectral and temporal diversity of the CSI, the proposed system can correctly extract the breathing rate traces of multiple users even if some of them merge together for a short time period. Furthermore, by utilizing the breathing traces obtained, the crowd number can be estimated for the occupancy level estimation in the smart home or smart office scenario.

In the second part, we propose a multi-person Respiration Rate (RR) as well as Heart Rate (HR) monitoring system leveraging the Channel Impulse Response (CIR) of a 60GHz WiFi. A calibration-free object detector is first designed to identify static objects, stationary human subjects and human in motion using both the amplitude and phase of the CIR measurement. To get robust HR estimations corresponding to stationary human subjects, the respiration signal is first eliminated from the phase of the CIR measurement before obtaining the spectrogram of heartbeat signal. Dynamic programming is further adopted to get the final estimation of HR by exploiting both the temporal and spectral information. Experimental results demonstrate promising performance of the proposed system, including the Non-Line-of-Sight (NLOS) scenario.

To further get finer information of heartbeat signal, in the third part, we propose mmHRV, the first multi-user Heart Rate Variability (HRV) monitoring system, using a commercial mmWave Frequency-Modulated-Continuous-Wave (FMCW) radar. We first develop a calibration-free target detector to identify the number of users and their locations. Then the heartbeat signal of each user is obtained by optimizing the decomposition of the composite phase measurement modulated by the chest movement. The exact time of heartbeats are estimated by identifying the peak location of the estimated heartbeat signal, and Inter-Beat Intervals (IBI) can be further derived to evaluate HRV. Extensive experiments have been conducted to explore the influence of different settings, including the distance between human and device, user orientation, incidental angle and NLOS setting, etc..

In the final part of this dissertation, we propose a driver vital sign monitoring system built upon a commercial FMCW radar. The system first eliminates driver's motion artifacts by a two-step motion compensation module. Then the respiration and heartbeat signals are estimated simultaneously by jointly decomposing the phase measurement over all range-azimuth bins containing vital signals. The RR, HR and IBI are further derived using the estimated respiration and heartbeat signals. We evaluate the system performance in real driving environment, where the impact of pavement condition, device location as well as motion type are explored in the experiment.

WIRELESS SENSING FOR VITAL SIGNS MONITORING

by

Fengyu Wang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:

Professor K. J. Ray Liu, Chair/Advisor

Professor Min Wu

Professor Furong Huang

Dr. Chenshu Wu

Professor Lawrence C. Washington, Dean's Representative

© Copyright by
Fengyu Wang
2021

Dedication

To my family—

Jiawang Wang, Ailan Luo and Guobin Wang

Acknowledgments

First and foremost, I would like to express my deepest appreciation to my advisor Prof. K. J. Ray Liu for providing me the opportunity to work on challenging and highly interesting projects over the last four years. I am extremely grateful to his valuable advice not only on my academic research but also for my life. His pursuit of excellence on academic, research and teaching not only leads me to the innovative research that contributes to the advance of wireless sensing, but also has a profound influence in my personal development as a role model.

I would like to thank all of my dissertation committee members, Prof. Min Wu, Prof. Furong Huang, Prof. Lawrence Washington and Dr. Chenshu Wu, for their time being my dissertation committee and reviewing my thesis. I would also like to thank Dr. Beibei Wang and Dr. Chenshu Wu for their countless suggestions on my research and patience during our discussions.

My sincere thanks also goes to all the group members of Signal and Information Group (SIG) for their support and help during these years. I thank Dr. Feng Zhang, Dr. Chen Chen and Dr. Qinyi Xu for their help and mentoring during my early years of PhD study. I would like to thank Dr. Xiaolu Zeng, Dr. Sai Deepika Regani, Yuqian Hu, Yusen Fan, Ruichen Wang and Muhammed Zahid Ozturk for the stimulating discussions. The collaboration with everyone is fantastic and I am very grateful to be a member of SIG family.

Last but not least, I would like to give my deepest appreciation to my parents, Jiwang Wang and Ailan Luo, thanks for your unconditional love and sup-

port. I thank my brother Guobin Wang, who is helping to take care of the whole family when I study abroad, thanks for your dedication. I owe my deepest thanks to my grandpa, Quanrong Wang, who gives me my name, and teach me how to be a good person. You set a good role model to me and my whole life will benefit from your guide. I feel so sorry that I was not able to keep accompany with you when you went away, I miss you so much. Words cannot express my gratitude to my family. This dissertation is dedicated to them.

Table of Contents

List of Tables	viii
List of Figures	ix
1 Introduction	1
1.1 Motivation	1
1.2 Related Works	3
1.2.1 Respiration Rate Estimation	4
1.2.2 Heart Rate Estimation	5
1.2.3 Heart Rate Variability Estimation	6
1.2.4 Handling Motion Artifacts	7
1.3 Dissertation Outline and Contributions	8
1.3.1 Respiration Tracking for People Counting (Chapter 3)	9
1.3.2 Multi-person Respiration Rate and Heart Rate Monitoring (Chapter 4)	10
1.3.3 mmHRV: Contactless Heart Rate Variability Monitoring (Chapter 5)	10
1.3.4 Driver Vital Signs Monitoring (Chapter 6)	11
2 Primer of Wireless Sensing	12
2.1 Wireless Channel Model for 2.4/5GHz WiFi	12
2.2 Wireless Channel Model for 60GHz WiFi	14
2.3 Wireless Channel Model for FMCW radar	17
3 Respiration Tracking for People Counting	20
3.1 Introduction	20
3.2 System Overview	22
3.3 Multi-user Breathing Spectrum Generation	24
3.4 Breathing Rate Trace Tracking	27
3.4.1 From Breathing Rates to Breathing Rate Traces	27
3.4.2 Extracting Breathing Rate Traces	29
3.4.2.1 Theoretical Model	29

3.4.2.2	Iterative Dynamic Programming	31
3.4.2.3	Detecting Empty Case	34
3.4.3	Trace Concatenating	35
3.5	People Counting	37
3.6	Experiments and Evaluation	39
3.6.1	Device and Methodology	40
3.6.2	Overall Performance	41
3.6.3	Performance Gain of Individual Modules	42
3.6.3.1	Impact of SNR Boosting Algorithm	43
3.6.3.2	Impact of IDP Estimation Algorithm	45
3.6.3.3	Impact of Quasi-bilateral Filter	45
3.6.4	Resolution Discussion	46
3.7	Summary	47
4	Multi-person Respiration Rate and Heart Rate Monitoring	49
4.1	Introduction	49
4.2	System Overview	51
4.3	Target Detection	52
4.3.1	Reflecting Objects Detector	52
4.3.2	Motion Detector	54
4.3.2.1	Static Reflecting Objects Detection	55
4.3.2.2	Stationary Human Subjects Detection	57
4.3.2.3	Motion Detection	58
4.3.3	Cell Merging/Clustering	59
4.4	Heart Rate Estimation	61
4.4.1	Breathing Interference Elimination	61
4.4.2	Heart Rate Estimation using Spectrogram	64
4.5	Experiment Evaluation	67
4.5.1	Overall performance	68
4.5.2	Impact of Distance	69
4.5.3	Impact of Orientation	71
4.5.4	Impact of Incident Angle	73
4.5.5	NLOS Case	75
4.5.6	Impact of User Heterogeneity	77
4.5.7	Impact of Body Movement	77
4.5.8	Multi-user Case	78
4.6	Summary	80
5	mmHRV: Contactless Heart Rate Variability Monitoring	82
5.1	Introduction	82
5.2	System Overview	83
5.3	Target Detection	84
5.3.1	Range-FFT and Digital Beamforming	85
5.3.2	Reflecting Object Detector	87
5.3.3	Human Subjects Detector	88

5.4	Heartbeat Extraction and HRV Estimation	89
5.4.1	Heartbeat Extraction Algorithm	90
5.4.1.1	Problem Formulation	90
5.4.1.2	Algorithm Design	93
5.4.2	HRV Estimation	96
5.5	Experiment Evaluation	97
5.5.1	Methodology	97
5.5.2	Overall Performance	98
5.5.3	Impact of Distance	102
5.5.4	Impact of Orientation	103
5.5.5	Impact of Incident Angle	104
5.5.6	LOS vs NLOS	105
5.5.7	Impact of User Heterogeneity	106
5.5.8	Multiple-User Case	107
5.6	Summary	109
6	Driver Vital Signs Monitoring	110
6.1	Introduction	110
6.2	System Overview	111
6.3	Vital Motion Extraction	113
6.3.1	Clutter Removal	113
6.3.2	Motion Compensation	114
6.3.2.1	Large Body Movement Compensation	115
6.3.2.2	Fine Movement Cancellation	117
6.3.3	Vital Bin Identification	118
6.4	Vital Signs Estimation	119
6.4.1	Minimization w.r.t. $u_{k,b}$	120
6.4.2	Minimization w.r.t. ω_k	121
6.4.3	Vital Signals Reconstruction	123
6.5	Experiment Evaluation	124
6.5.1	Methodology	125
6.5.2	Overall Performance	126
6.5.3	Impact of Road Condition	127
6.5.4	Impact of Device Location	129
6.5.5	Impact of Motion Type	131
6.5.6	Impact of User Heterogeneity	132
6.6	Summary	133
7	Conclusions and Future Work	135
7.1	Conclusions	135
7.2	Future Work	137
	Bibliography	139

List of Tables

3.1	Accuracy with different frequency and spatial separation - Part I. .	47
3.2	Accuracy with different frequency and spatial separation - Part II.	48
4.1	Performance with different blockage materials.	76
4.2	Impact of thickness of clothes.	76
4.3	Performance for different motion states.	78
4.4	Performance for different separation angles.	80
5.1	HRV estimation results in terms of mean IBI, RMSSD, SDRR and pNN50 for 11 subjects (part-I).	100
5.2	HRV estimation results in terms of mean IBI, RMSSD, SDRR and pNN50 for 11 subjects (part-II).	101

List of Figures

2.1	Coordinate system and typical signal containing vital sign.	15
2.2	Example of the frame structure and the cell.	17
2.3	FMCW radar system.	18
3.1	Processing flow of system.	23
3.2	PSD of different links. Ground-truth: 3 people sitting in car with breathing rate [10 14 15] Respiration Per Minute (RPM) (marked as dashed lines).	25
3.3	Normalized PSD of different links after SC combination.	26
3.4	PSD after link combination.	27
3.5	Spectrogram after link combination.	29
3.6	Successive cancellation procedure of IDP.	34
3.7	Traces found by IDP in four adjacent time windows.	36
3.8	Trace concatenating result of windows in Fig. 3.7.	37
3.9	Example and illustration of quasi-bilateral filter.	39
3.10	Experiment setup.	41
3.11	Confusion matrix of people counting in LAB.	42
3.12	Confusion matrix of people counting in car.	43
3.13	TP comparison of different algorithms.	43
3.14	Accuracy of different algorithms.	44
3.15	Experiment setup for resolution investigation.	46
4.1	An overview of ViMo.	50
4.2	CFAR window.	53
4.3	Example of 1D-CFAR.	55
4.4	Example phase and ACF for target detection.	56
4.5	Example of cell merging.	59
4.6	Example of breathing interference elimination.	62
4.7	Spectrogram of residual signal.	64
4.8	Experiment setup.	68
4.9	Experiment setup and result for the impact of distance.	70
4.10	Experiment setup and result for the impact of orientation.	72
4.11	Experiment setup and result for the impact of incident angle.	73

4.12	Experiment setup and result for the impact of blockage.	74
4.13	Experiment setup with different blockage materials.	75
4.14	Impact of user heterogeneity on estimation accuracy.	77
4.15	Experiment setup for the impact of separation angle.	79
4.16	Multi-user experiment setup.	81
4.17	Multi-user accuracy and detection performance.	81
5.1	Processing flow of mmHRV.	84
5.2	Antenna Deployment.	85
5.3	Example of the reflecting object detector.	86
5.4	Example of the human subject detector.	89
5.5	Example of heartbeat extractor.	92
5.6	Example of IBI estimation.	94
5.7	Experiment setup.	98
5.8	Over all performance of the IBI estimation error.	99
5.9	Experiment setup and the absolute IBI estimation results versus distance.	103
5.10	Experiment setup and the absolute IBI estimation results versus orientation.	104
5.11	Experiment setup and the absolute IBI estimation results versus incident angle.	105
5.12	Experiment setup and the absolute IBI estimation results versus blockage.	106
5.13	Impact of user heterogeneity.	107
5.14	Mean and RMSE of the absolute IBI estimation error of multiple users.	108
6.1	Processing flow of system.	111
6.2	Example of clutter removal.	114
6.3	Example of consecutive frame after clutter removal.	115
6.4	Example of large body movement compensation.	116
6.5	Example of target detection.	117
6.6	Example of fine movement cancellation.	118
6.7	Example of phase decomposition of 3 vital bins.	122
6.8	Example of estimated result v.s. ground truth.	123
6.9	Experiment setup.	124
6.10	Experiment path.	125
6.11	Comparison of vital sign estimation performance between proposed method and V ² iFi.	127
6.12	Vital sign estimation performance versus pavement condition.	128
6.13	Vital sign estimation performance versus device location.	130
6.14	Vital sign estimation performance versus motion type.	131
6.15	Impact of user heterogeneity.	133

List of Abbreviations

ACF	Autocorrelation Function
ADAS	Advanced Driver Assistance System
ADC	Analog-to-Digital Converter
ANS	Autonomic Nervous System
AoA	Angle of Arrival
BPF	Band-Pass Filter
BPFB	Band-Pass Filter Bank
BPM	Beat Per Minute
CDF	Cumulative Distribution Function
CFAR	Constant False Alarm Rate
CIR	Channel Impulse Response
COTS	Commodity Off-The-Shelf
CSI	Channel State Information
CUT	Cell Under Test
DI	Detection Index
DP	Dynamic Programming
ECG	Electrocardiogram
EEMD	Ensemble Empirical Mode Decomposition
EM	Electromagnetic
FFT	Fast Fourier Transform
FI	False-alarm Index
FM	Frequency Modulation
FMCW	Frequency Modulated Continuous Wave
FoV	Field of View
HPF	High-Pass Filter
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
IDP	Iterative Dynamic Programming
I.I.D.	Independent and Identical Distribution
IMF	Intrinsic Mode Function
IoT	Internet of Things

LOS	Line of Sight
MAE	Median Absolute Error
MIMO	Multiple Input Multiple Output
mmWave	Millimeter Wave
MPC	Multipath Component
MS-VMD	Multi-Sequence Variational Mode Decomposition
NLOS	Non Line of Sight
OFDM	Orthogonal Frequency-Division Multiplexing
PDF	Probability Density Function
PPG	Photoplethysmogram
PSD	Power Spectrum Density
RBM	Random Body Motion
RF	Radio Frequency
RMSE	Root Mean Square Error
RMSSD	Root Mean Square of Successive Differences
RPM	Respiration Per Minute
RR	Respiration Rate
RSS	Received Signal Strength
Rx	Receiver
SC	Subcarrier
SINR	Signal-to-Interference-plus-Noise-Ratio
SNR	Signal-to-Noise-Ratio
STFT	Short Term Fourier Transform
TP	True Positive
Tx	Transmitter
UWB	Ultra-Wide Band

Chapter 1: Introduction

1.1 Motivation

In recent years, human health monitoring technologies have gained more and more attention to help people access their daily health status. To live a healthy lifestyle, it is critical to keep track of vital signs regularly and frequently. Besides, the current COVID-19 pandemic has placed new demands on the health monitoring system. Note that the non-critical patients are mostly referred to perform quarantine at home due to the limited capacity and resources of current health care system, it is important to support the vital signs monitoring at home.

Traditional vital signs monitoring requires medical professionals using dedicated equipment, so it is usually performed in the hospital at a very limited frequency. Thanks to the advancements in on-body sensors, people can monitor their health status much easier using wearable devices. However, the sensor-based methods require physical contact with human body, which is uncomfortable to use for a long period of time and may cause skin allergies. Moreover, these devices suffer from high cost and inconvenience in sharing with other people. Therefore, the technologies for non-intrusive vital signs monitoring have gained more and more attention. As a less intrusive solution, vision-based methods uti-

lize image sequences to detect the vital signs. However, the main drawbacks such as its poor performance in low-light scenarios and the privacy concerns hinder the wide deployment of the vision-based systems.

Thanks to the ubiquitous deployment of wireless radio devices and the development of wireless sensing [13, 36, 73], Radio Frequency (RF) based methods have become one of the most promising candidates. Intuitively, the presence of a human subject will affect the RF propagation [84, 89, 90], i.e., RF signals reflected off human subjects will be modulated by the body movement including chest movement due to respiration and heartbeat. As a result, RF-based systems can estimate vital signs without any physical contact, while preserving the user privacy and operating robustly regardless of the light conditions.

In the meanwhile, vital signs monitoring in driving environments is another paramount application. As automobiles have become an essential part to facilitate our daily life, Advanced Driver Assistance Systems (ADAS) have been gaining more and more interest in assisting drivers to enhance both safety and convenience. To respond timely in case of an emergency, ADAS needs to keep track of the driver's health/consciousness, which is generally achieved by monitoring driver's vital signs including Respiration Rate (RR), Heart Rate (HR) and Heart Rate Variability (HRV). The RF-based vital signs monitoring systems attach great importance due to the superiority in handling environment change (e.g., light and temperature condition) in driving scenario while providing a non-intrusive solution to detect driver's vital signs.

Motivated by the aforementioned problems in current indoor and in-car

vital signs monitoring applications, in this dissertation, we illustrate how wireless sensing can be used to achieve high-accuracy vital sign monitoring for smart home and smart car scenarios.

1.2 Related Works

The past few decades have witnessed a surging demand of vital signs monitoring systems. Compared with the traditional method that requires users to wear contact sensors, the contactless method can alleviate users' burden and reduce the device cost. Existing approaches for contactless vital signs monitoring can be classified into two categories: vision-based method and RF-based method.

Vision-based methods utilize the image sequences to monitor vital signs. In principle, the breathing process causes involuntary quasi-periodic thoracic and abdominal movements, which can be captured by video stream and thus utilized for respiration rate estimation [67]. It has also been studied that the skin color changes caused by blood perfusion can be used as a good feature to estimate HR [32, 95] and HRV [38, 41, 43, 51, 62]. However, the main drawbacks such as the sensitivity to the light conditions, the Line-of-Sight (LOS) requirement as well as the privacy invasion hinder the wide deployment of the vision-based systems.

To this end, RF-based sensing has become one of the most promising candidates due to the proliferation of wireless devices. Since the presence of a human subject will affect propagation of electromagnetic (EM) wave [36, 89, 90], the RF signals reflected off human body will be modulated by the body movement (e.g.,

periodic chest movement caused by respiration and heartbeat). As a result, vital information of the human subject can be unveiled by analyzing the channel propagation characteristics [13, 36, 73, 76]. This dissertation is widely related to RF-based vital signs monitoring, and we will review the related works in the following subsections.

1.2.1 Respiration Rate Estimation

Respiration is an important vital indicator of health status and medical diagnosis. RF-based monitoring solutions can estimate breathing rates contactlessly since chest motions can be captured by RF signals. Due to the availability of the Received Signal Strength (RSS) measurement on most WiFi devices, UbiBreathe [3] is proposed to estimate respiration rate using RSS. However, since RSS is not sensitive to the minute chest movements, the setting should be well designed to get a good accuracy. Compared to RSS, Channel State Information (CSI) is a fine-grained information that can portrait the EM wave propagation and is more sensitive to the minute chest movement [11, 12, 82, 87, 88, 91]. However, due to the omni-directional propagation and the narrow bandwidth, it is impossible to isolate each individual's breathing signal. So most of the previous works either study single person scenario [82, 87, 91] or assume the breathing rates of each individuals are distinct [11, 12, 88].

1.2.2 Heart Rate Estimation

Since the perturbation caused by the heartbeat (0.2 ~ 0.5 mm) [57] is much smaller than the wavelength of 2.4/5GHz WiFi system (60 ~ 120 mm), the phase change caused by the heartbeat is very small, resulting in a low SNR. Thus, most of the conventional WiFi-based systems cannot estimate heart rates. To overcome the coarse range as well as spatial resolution limitation in WiFi-based systems, researchers try to build dedicated radar systems to remotely monitor vital signs. Ultra-wideband (UWB) [29, 58, 66] and Doppler radar [21, 22, 31, 37, 44, 47, 50, 54, 56, 77] either directly measure the distance between chest and device or the relative speed change of chest movement to get the estimation of chest displacement caused by vital signs. However, the assumption of a single user in these works limits the further deployment of the system. Frequency-Modulated-Continuous-Waves (FMCW) radar is built in [5] [45] to measure both the respiration rate (RR) and heart rate (HR). Leveraging the fact that different users may locate in distinct bins (a.k.a, range buckets), the vital signs of multiple people can be monitored simultaneously. Finer spatial resolution is achieved in mmVital [85] by using a pair of horn antennas, which investigates the feasibility of using 60GHz mmWave signal to simultaneously monitor vital signs in a multi-user case. However, most of these works [5, 21, 22, 29, 45, 58, 85] try to directly utilize frequency analysis and Band-Pass Filter (BPF) to estimate the heart rate. As a result, these methods are easy to fail when the subject's heart rate is close to the respiration harmonics. The polynomial fitting has been used [77] [44] [50] to remove respiration

motion. However, the order of polynomial needs to be carefully selected by empirical experience, and under-fitting or over-fitting can be easily triggered when the experimental setting is changed (e.g., change of sampling rate or window length).

1.2.3 Heart Rate Variability Estimation

Dedicated radars have been used to measure the distance change between the chest and the device to estimate HR, as discussed in Section-1.2.2. However, those systems mainly rely on the frequency-domain spectral analysis to estimate the HR, which may take a couple of seconds, making it impossible to estimate the precise timing of each heartbeat for calculating the HRV.

To achieve robust HRV estimation, a preliminary work [63] tries to eliminate the respiration effect by asking users to hold their breaths. However, holding breath will impact the performance of the HRV and thus the HRV metrics estimated in this condition cannot indicate the users' health condition accurately. To extract the heartbeat wave, the 2nd-derivative of the distance change (i.e., acceleration) has been considered in [93], which is equivalent to a High-Pass Filter (HPF). However, the residual signal is too noisy for heartbeat extraction. To identify the exact time of each heartbeat, it is assumed that the heartbeat signal is the successive multiple copies of a heartbeat template with different time scales. However, as shown in [80], using a single template is insufficient. To get a high detection accuracy, multiple templates are needed, and training is required

before estimation. A simple approach based on two Band-Pass Filters (BPFs) is employed for separating the respiration and the heartbeat signal in [24] [49], however, since undesired peaks caused by the interference (e.g., harmonics of respiration) may overlap with the dominant frequency band of the heartbeat, dedicated systems need to be employed to remove the false peaks. Ensemble Empirical Mode Decomposition (EEMD) is applied in [24] to remove the noise and interference, however, the mode mixing problem as well as selecting the optimal decomposed Intrinsic Mode Functions (IMFs) is not easy to solve in real applications. IBIs are assumed to not change much in [49], and the auto-correlation is used to remove the false peaks. However, the system can only perform well when the passing band does not overlap with the strong interference. To reduce the interference, the Band-Pass Filter Bank (BPFB) is applied in [52], where the HR is first estimated and the heartbeat signal is then filtered by using the BPF with a center frequency at HR. However, a large error will occur once the HR is not estimated correctly.

1.2.4 Handling Motion Artifacts

Preliminary works [5, 11, 45, 49, 50, 52, 75, 76, 79, 85, 93] have shown the feasibility to monitor RR, HR and HRV using RF signals. However, these systems can merely work in the ideal indoor scenario with the hypotheses of stationary human subject, and cannot directly apply to driver vital sign monitoring scenario due to the frequent driving motion. Note that even subtle displacement of body

roaming could be larger than the displacement caused by respiration and heart-beat. As a result, handling the motion artifacts is one of the biggest challenges to achieve RF-based driver's vital sign monitoring.

Multiple transceivers are deployed at opposite sides of human body to eliminate body movement [33] [70]. However, it significantly increases the system complexity and deployment cost, and thus making it hard to implement in practice. The correlation of range taps between different time blocks is used in [65] [30] to remove the body movement in the system with a single transceiver. However, this method can only remove specific body movement that is larger than the range resolution, and the motion artifacts within the range resolution still remains, thus reducing the estimation accuracy. To remove the motion artifacts located in the same range tap, polynomial fitting is used in [86] to estimate the displacement caused by body motion. However, the order of the polynomial fitting needs to be carefully selected for different motion types, which is not robust in practical use. Note that the above systems can only estimate HR, and the residual signal after motion elimination is too noisy to extract the exact time of heartbeats for further HRV estimation.

1.3 Dissertation Outline and Contributions

From the previous discussion, we can see that the wireless sensing is a promising candidate for non-intrusive vital signs monitoring for both smart home and smart car applications. However, there are many challenges to be solved to

achieve accurate vital signs detection. In this dissertation, we focus on leveraging the commercial wireless devices to accurately detect the vital signs using wireless sensing techniques. The rest of this dissertation is organized as follows.

1.3.1 Respiration Tracking for People Counting (Chapter 3)

In this chapter, we propose a respiration tracking system using the Channel State Information (CSI) of a single pair of commercial WiFi devices. The system is capable of continuously tracking the breathing rates of multiple persons without any prior knowledge of crowd number. Considering that the breathing signals are typically fairly weak on a single subcarrier, we make full use of the multiple subcarriers and antenna links in Multiple-Input and Multiple-Output Orthogonal Frequency-Division Multiplexing (MIMO-OFDM) system and propose an adaptive subcarrier combining method to boost the Signal-to-Noise-Ratio (SNR) of breathing signals. To track the breathing rate traces from the spectrogram, a Markov Chain Model is introduced to handle dynamics in natural breathing, and we propose a successive cancellation scheme that resolves each individual's breathing trace one by one. Furthermore, by utilizing the breathing traces obtained, the system can estimate the occupancy level in quasi-static scenarios. We prototype and evaluate the proposed system on Commodity Off-The-Shelf (COTS) WiFi devices. The results demonstrate promising performance.

1.3.2 Multi-person Respiration Rate and Heart Rate Monitoring (Chapter 4)

In this chapter, we propose *ViMo*, a calibration-free non-contact **Vital sign Monitoring** system that can detect stationary/non-stationary users and estimate the Respiration Rates (RRs) as well as Heart Rates (HRs) built upon a commercial 60GHz WiFi. The system consists of two key components. First, we design a calibration-free object detector that can identify static objects, stationary human subjects and human in motion. Second, to get robust HR estimation, we eliminate the respiration signal from the phase of the Channel Impulse Response (CIR) by smoothing spline algorithm. Dynamic Programming (DP) is applied to further resist the random measurement noise. The influence of different settings, including the distance between human and the device, user orientation and incidental angle, blockage material, body movement and conditions of multi-user separation are investigated by extensive experiments. Experimental results show that ViMo monitors user's vital signs accurately, with a median error of 0.19 Respiration Per Minute (RPM) and 0.92 Beat Per Minute (BPM).

1.3.3 mmHRV: Contactless Heart Rate Variability Monitoring (Chapter 5)

In this chapter, we propose mmHRV [74], the first contact-free multi-user Heart Rate Variability (HRV) monitoring system using commercial mmWave ra-

dio. Different from Chapter 4, in this chapter, finer information of heartbeat is extracted. To get the exact time of heartbeats, we optimize the decomposition of the phase of the channel information modulated by the chest movement. Extensive experiments have been conducted to evaluate the performance. It has been shown that the median error of Inter-Beat Intervals (IBI) is 28ms (w.r.t. 96.16% accuracy). The performance of the multi-user scenario is slightly degraded compared with the single-user case, however, the median error of the 3-user case is within 52ms for all 3 tested locations.

1.3.4 Driver Vital Signs Monitoring (Chapter 6)

In this chapter, we propose a robust driver's vital sign monitoring system using commercial mmWave radio. To extract the reflection signals containing vital signals, the motion artifacts are first removed by a novel motion compensation module. The bins containing vital signals are then identified by checking the periodicity of the phase measurement. The respiration and heartbeat wave are further reconstructed by jointly optimizing the decomposition of all the extracted compound vital signals. We evaluate the system performance in real driving environment and investigate the impact of different parameters including the device locations, pavement conditions and motion types. Experimental results show that the proposed system can achieve a median error of 0.16 RPM, 0.82 BPM and 46 ms for RR, HR and IBI estimations, corresponding to the relative accuracy of 99.17%, 98.94% and 94.11%, respectively.

Chapter 2: Primer of Wireless Sensing

Wireless channel information plays an important role in wireless sensing. Since the presence of human subject affects the wireless signal propagation, the information corresponding to the human subject is embedded in the channel information. In this chapter, we give a detailed introduction to the wireless channel model and illustrate why channel information can depict the vital motions for wireless vital signs monitoring. The rest of this chapter is organized as follows. The Channel State Information (CSI) of 2.4/5GHz WiFi is introduced in Section 2.1. We further introduce the Channel Impulse Response (CIR) of 60GHz WiFi and Frequency-Modulated-Continuous-Wave (FMCW) radar in Section 2.2 and Section 2.3 respectively.

2.1 Wireless Channel Model for 2.4/5GHz WiFi

Given a typical transmission pair of WiFi devices equipped with omnidirectional antennas, the transmitted signal encounters different scatters in the environment and thus reaches the receiving antenna through different paths. Therefore, the received signal is a superposition of multiple scaled and delayed replicas of the transmitted signal. The Channel State Information (CSI) is the frequency

response of the channel, which can be modeled as

$$h(t, f_k) = \sum_{l=1}^L a_l(t) \exp(-j2\pi f_k \tau_l(t)), \quad (2.1)$$

where $k \in \mathcal{V}$ is the subcarrier (SC) index with center frequency f_k in the set of usable SCs \mathcal{V} . L is the total number of multipath components (MPCs), while $a_l(t)$ denotes the complex gain of MPC l . The propagation delay $\tau_l(t)$ is a function of the propagation distance: $\tau_l(t) = \frac{d_l(t)}{c}$, where c is the speed of light.

In the presence of human beings, one or more paths of signal propagation will be altered due to the human motion. In the presence of breathing, (2.1) can be rewritten as

$$h(t, f_k) = \sum_{i \in \mathcal{I}} \sum_{l \in \Omega_{d_i}} a_l(t) \exp(-j2\pi f_k \frac{d_l(t)}{c}) + \sum_{l \in \Omega_s} a_l \exp(-j2\pi f_k \frac{d_l}{c}), \quad (2.2)$$

where \mathcal{I} denotes the set of human subjects. Ω_{d_i} denotes the MPCs scattered by human being i , resulting in time-variant complex gain and delay. Ω_s denotes the MPCs that are not affected by people's breathing, whose complex gain and delay keep time-invariant. It is noted that for each MPC subset Ω_{d_i} , the delay is periodic due to the periodic chest movement, i.e., $d_l(t + T_{b_i}) = d_l(t), \forall l \in \Omega_{d_i}$, where T_{b_i} is the breathing cycle of i -th human subject. Hence we would be able to see multiple frequency components of the measured CSI, each corresponding to a distinct breathing signal. The respiration rate traces corresponding to different human subjects are estimated from the measured CSI time series in Chapter 3.

2.2 Wireless Channel Model for 60GHz WiFi

60GHz WiFi is becoming a mainstream in wireless devices to enable high rate networking with the development of 802.11ad/ay standards (*a.k.a.* WiGig). 60GHz WiFi offers high directionality with large phased arrays in small size thanks to millimeter-wavelength and precise time-of-flight measurements brought by the large bandwidth. In the following, we will discuss the impact of vital signals on the Channel Impulse Response (CIR) of the Qualcomm's 60GHz WiFi.

Assume the travelling distance of the electromagnetic (EM) wave reflected by human chest is $d(t)$, then the CIR between Tx antenna m and Rx antenna n can be expressed as

$$h_{m,n}(t) = a_{m,n}(t) \exp(-j2\pi \frac{d_{m,n}(t)}{\lambda_c}), \quad (2.3)$$

where $a_{m,n}(t)$ is the complex channel gain, λ_c denotes the wavelength of the carrier. Due to the modulation of the vital signs, i.e., respiration and heartbeat, $d_{m,n}(t)$ appears to be a combination of two periodic signals, which can be further expressed as

$$d_{m,n}(t) = d_0(m, n) + s_r(t) + s_h(t), \quad (2.4)$$

where $s_r(t)$ and $s_h(t)$ denote the distance change due to respiration and heartbeat, and $d_0(m, n)$ denotes the constant travelling distance. Considering the fast-time resolution of the device, the reflected signal will fall into the l -th tap if $d_0(m, n) = lT_s c + \Delta d(m, n)$, where $T_s = 1/B$ denotes the fast-time resolution and B stands for the system bandwidth. $\Delta d(m, n)$ denotes the distance residual, where $\Delta d(m, n) =$

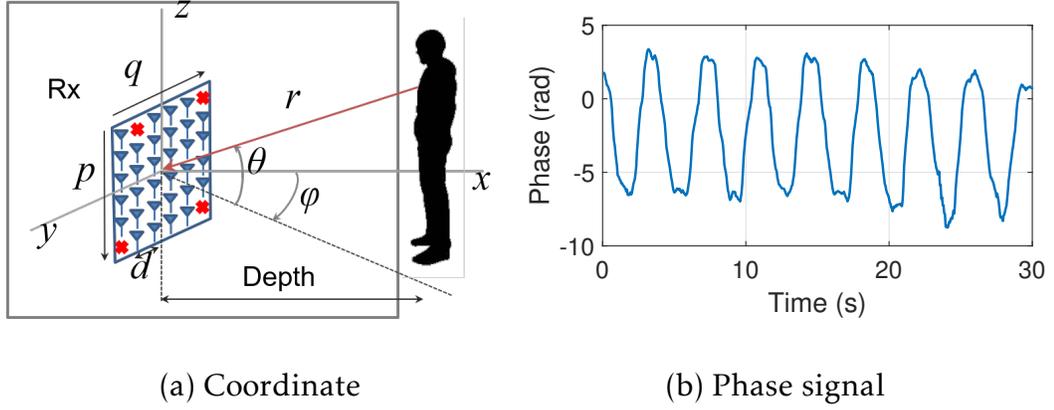


Figure 2.1: Coordinate system and typical signal containing vital sign.

$$d_0(m, n) - lT_s c < T_s c.$$

If the reflected signal falls into the l -th tap of the measured CIR with residual $\Delta d(m, n)$, the CIR at tap l , denoting as $\mathbf{h}_l(t) = [h_{1,1}(t), h_{1,2}(t), \dots, h_{M,N}(t)]^T$, can be expressed as

$$\begin{aligned} \mathbf{h}_l(t) &= \mathbf{a}(t) \odot \exp(-j2\pi \frac{\Delta \mathbf{d} + s_r(t) + s_h(t)}{\lambda_c}) \\ &= \tilde{\mathbf{a}} \exp(-j2\pi \frac{s_r(t) + s_h(t)}{\lambda_c}), \end{aligned} \quad (2.5)$$

where $\Delta \mathbf{d} = [\Delta d(1, 1), \Delta d(1, 2), \dots, \Delta d(M, N)]^T$, $\mathbf{a}(t) = [a_{1,1}(t), a_{1,2}(t), \dots, a_{M,N}(t)]^T$, and \odot denotes elementwise product. We assume $\mathbf{a}(t)$ is time-invariant due to the tiny movement of the subject, and the common phase shift is absorbed in the term $\tilde{\mathbf{a}}$. It is apparent that the phase of the CIR measurement changes periodically in slow time due to the periodic motions of respiration and heartbeat, as shown in (2.5). Fig. 2.1 (b) shows a typical phase signal containing vital signs.

To further get the CIR at a specific angle, beamforming can be performed at both Tx and Rx. Here we take the conventional beamforming as an example. In specific, at Tx side, to create a beam toward to angle (θ, ϕ) , the coefficient of the

m -th antenna of steering vector \mathbf{s}_{Tx} is

$$s_{\text{Tx},m}(\theta, \phi) = \exp\left(-j2\pi \frac{d_{m,x} \cos \theta \sin \phi + d_{m,y} \cos \theta \cos \phi}{\lambda_c}\right). \quad (2.6)$$

$d_{m,x}$ and $d_{m,y}$ are the horizontal and vertical distance between antenna m and the origin, as shown in Fig. 2.1 (a). Similarly, at Rx side, to detect the Angle of Arrival (AoA) of the signal, the coefficient of the n -th antenna of steering vector \mathbf{s}_{Rx} is

$$s_{\text{Rx},n}(\theta, \phi) = \exp\left(-j2\pi \frac{d_{n,x} \cos \theta \sin \phi + d_{n,y} \cos \theta \cos \phi}{\lambda_c}\right). \quad (2.7)$$

$d_{n,x}$ and $d_{n,y}$ are the horizontal and vertical distance between antenna n and the origin. The CIR after performing beamforming can be expressed as

$$h_{\theta,\phi,l}(t) = \mathbf{s}^H(\theta, \phi) \mathbf{h}_l(t) + \epsilon(t), \quad (2.8)$$

where $\epsilon(t)$ stands for additive white Gaussian noise which is independent and identically distributed (I.I.D) for different links. $\mathbf{s}^H(\theta, \phi)$ is the steering vector pointing to the direction (θ, ϕ) , which can be expressed as the Kronecker product between \mathbf{s}_{Rx} and \mathbf{s}_{Tx} , i.e.,

$$\mathbf{s}(\theta, \phi) = \mathbf{s}_{\text{Tx}}(\theta, \phi) \otimes \mathbf{s}_{\text{Rx}}(\theta, \phi). \quad (2.9)$$

The Qualcomm's 60GHz WiFi enables on-chip beamforming at both Tx and Rx. The frame structure is shown in Fig. 2.2 (a). In each burst, Tx will send a series of known pulses toward to different sectors (θ, ϕ) , and Rx will also perform beamforming with corresponding steering vector towards to angle (θ, ϕ) . Therefore, the space can be separated as cells, as shown in Fig. 2.2 (b), where $\Delta r = \frac{c}{2B}$ is the range resolution of the device, and l is the range tap index. $\Delta\theta$ and $\Delta\phi$

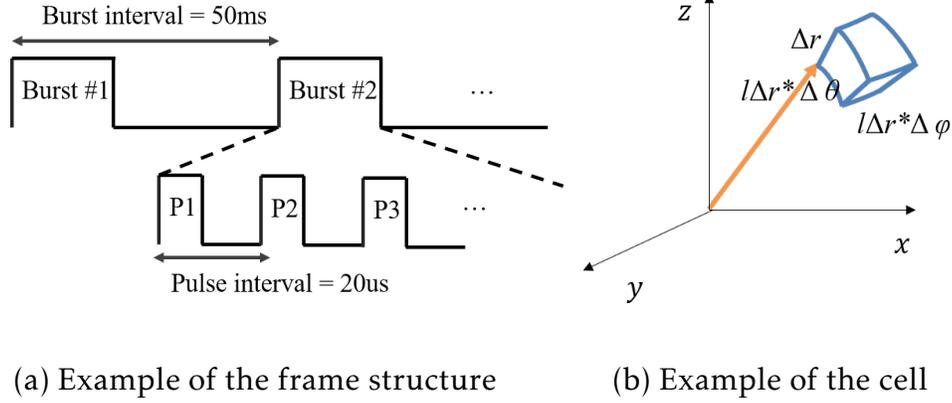


Figure 2.2: Example of the frame structure and the cell.

denote the elevation and azimuth separation respectively. Thus, in each burst, we will get channel information of all cells.

2.3 Wireless Channel Model for FMCW radar

With the development of the integrated circuit, it is possible to integrate a complete radar system on a single chipset. The dramatic decrease of the device cost makes radar systems become popular for industrial applications, where Frequency-Modulated-Continuous-Wave (FMCW) radar is widely adopted due to its accurate and stable measurement.

During the measurement, a chirp signal is transmitted by the FMCW radar, where the instantaneous transmitting frequency is a periodic linearly-increasing signal as shown in Fig. 2.3, and it can be expressed as [45]

$$f_t = f_c + \frac{B}{T_c}t, \quad (2.10)$$

where f_c is the chirp starting frequency, T_c is the chirp duration and B is the

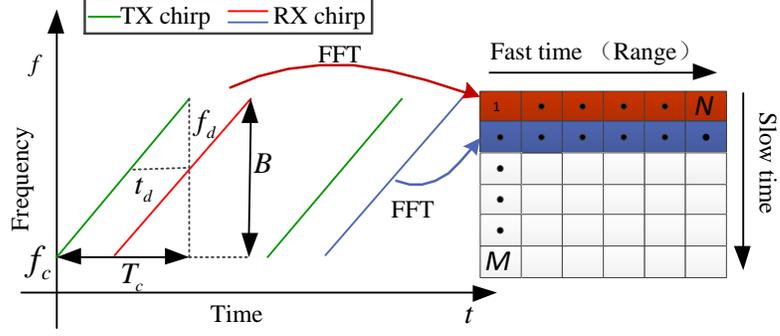


Figure 2.3: FMCW radar system.

bandwidth. According to Frequency Modulation (FM), the transmitted signal $x_T(t)$ can be expressed as

$$\begin{aligned}
 x_T(t) &= A_T \exp\{-j[2\pi \int_0^t f_t(\tau) d\tau]\} \\
 &= A_T \exp\{-j[2\pi f_c t + \pi \frac{B}{T_c} t^2]\},
 \end{aligned} \tag{2.11}$$

where A_T is the transmitting power. When the EM wave is reflected by human chest at distance $d(t)$, the reflected signal $x_R(t)$ can be expressed as

$$x_R(t) = A_R \exp\{-j[2\pi f_c(t - t_d) + \pi \frac{B}{T_c} (t - t_d)^2]\}, \tag{2.12}$$

where A_R is the amplitude of the receiving signal. t_d stands for the round-trip delay and can be denoted as $t_d = \frac{2d(t)}{c}$, where c is the speed of light.

Mixing the received signal with a replica of the transmitted signal and following a low-pass filter, the channel information $h(t)$ can be expressed as

$$h(t) = A \exp\{-j(2\pi \frac{Bt_d}{T_c} t + 2\pi f_c t_d - \pi \frac{B}{T_c} t_d^2)\}. \tag{2.13}$$

Note that the term $\pi \frac{B}{T_c} t_d^2$ is negligible, especially in short-range scenarios. There-

fore, the $h(t)$ can be written as

$$h(t) = A \exp \left\{ -j \left(2\pi \frac{Bt_d}{T_c} t + 2\pi f_c t_d \right) \right\}, \quad (2.14)$$

which is a sinusoidal signal whose frequency $f_b \triangleq \frac{Bt_d}{T_c} = \frac{2Bd(t)}{cT_c}$ depends on the target's distance. For each chirp, the baseband signal $h(t)$ is digitized by Analog-to-Digital Converter (ADC), producing N samples per chip, referred to as fast time. The time corresponding to the transmission of chirps is referred to as slow time, as shown in Fig. 2.3. Therefore, the digitized channel information for the n^{th} ADC sample and m^{th} chirp can be expressed as

$$h(n, m) = A \exp \left\{ -j \left(2\pi f_b n T_f + \frac{4\pi d(m T_s)}{\lambda_c} \right) \right\}, \quad (2.15)$$

where T_f and T_s are the time interval in fast time and slow time respectively. λ_c denotes the wavelength of the chirp. Therefore, when the signal is reflected by the human chest, the periodic change corresponding to respiration and heartbeat can be observed in the phase of the channel information, as shown in (2.15).

Chapter 3: Respiration Tracking for People Counting

3.1 Introduction

Human-centric sensing via wireless Radio Frequency (RF) has attracted an increasing interest for a range of Internet of Things (IoT) applications [73] [36]. Demands of accurate and passive awareness of the environment surge for many applications [13] [27]. For instance, a smart home can adjust the light and ventilation system based on occupancy level to improve energy efficiency [15] [14]. People recognition in smart homes enables user authentication for home security and privacy protection [7,78,83,92,96]. Besides understanding the environment, monitoring the status of human in the environment also has received great attention, and respiration/breathing rate, serving as a significant vital sign, has been an important topic for RF sensing.

Comparing with conventional methods that use dedicated sensors to monitor breathing rates, RF sensing provides contact-free solutions. Thanks to the easy-deployment property, WiFi-based methods have been studied in the past decade [3,11,12,35,88,91]. Received Signal Strength (RSS) measured by a WiFi device has been used in [3] to measure the chest movement during breathing. However, the accuracy of respiration rate estimation degrades when the test sub-

jects do not hold the device. Fine-grained Channel State Information (CSI) is more sensitive to the environment changes, which has been utilized to capture minute movements caused by respiration in [11, 12, 35, 88, 91]. However, due to the omni-directional propagation and narrow bandwidth of commonly used 2.4/5GHz WiFi, the received signal can be reflected from multiple humans in an indoor space. This makes it difficult to extract the vital signs of multiple humans from the reflected signal. Most of the previous works assume that there is only one person in the observation area [91] or assume the respiration rates of different people are distinct and the number of people is known in advance [11, 12, 35, 88].

In this chapter, we propose a solution to continuously track human respiration rate without any prior knowledge of the crowd number or assuming that the breathing rates of different users are distinct. Different from the previous works, we are particularly interested in matching the breathing rates estimated in different time instances to different users, i.e., which breathing rate corresponds to which person. By utilizing the estimated breathing rate traces, our system can achieve people counting at the same time. The rest of this chapter is organized as follows. The system is overviewed in Section 3.2, followed by multi-person breathing spectrum generation in Section 3.3 and breathing rates tracking in Section 3.4. The case of people counting is studied in Section 3.5 and the performance is evaluated in Section 3.6. Finally, conclusions are drawn in Section 3.7.

3.2 System Overview

The computational pipeline underlying the proposed system is shown in Fig. 3.1. Different from many previous works aiming at estimating independent breathing rates at certain time instances, this work focuses on utilizing the frequency as well as time domain information to do identity matching. The core idea is to estimate the breathing rate sequences along the time (a.k.a, breathing rate traces) of different individuals. Furthermore, utilizing the estimated breathing rate traces, we can estimate the occupancy level in the observation area. This idea immediately leads to three stages of the proposed system: **(1) Multi-user breathing spectrum generation**, **(2) breathing rate trace tracking**, and **(3) people counting**.

In the first stage, the proposed system first performs Short-Term Fourier Transform (STFT) on CSI measurements to extract the periodic breathing signals. As long as the breathing rates of different individuals are different, multiple frequency components would be observed in the frequency response. The extracted breathing signals are typically fairly weak on a single subcarrier, which are further boosted by a novel adaptive subcarrier combining method. Stage 1 finally outputs a spectrogram of the estimated breathing rates over time.

In Stage 2, the goal is to track the breathing rate traces (i.e., breathing sources) from the spectrogram obtained from Stage 1. However, there is a significant gap between breathing rates to breathing rate traces because of two reasons: First, different individuals may have the same breathing rates that overlap

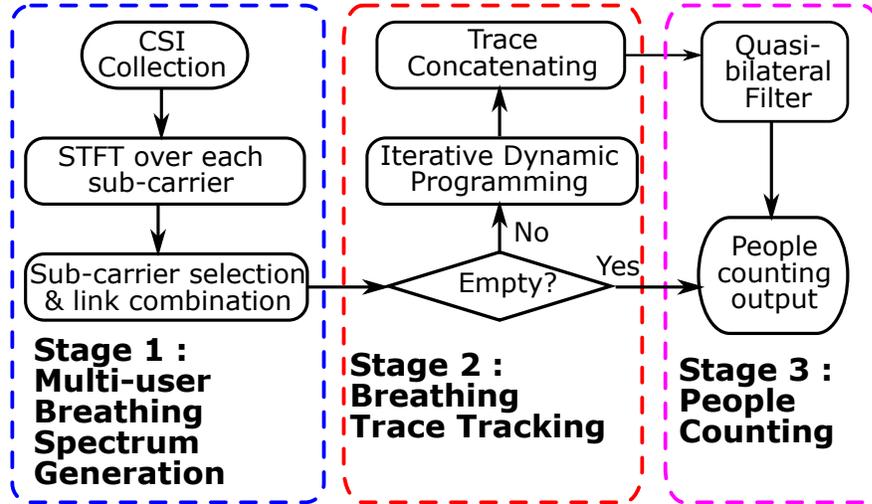


Figure 3.1: Processing flow of system.

with each other. Second, one’s breathing rate varies over time. To address the challenges, a Markov Chain Model is introduced to handle dynamics in natural breathing. We propose a successive cancellation scheme that resolves each individual’s breathing trace one by one by a novel algorithm of iterative dynamic programming. Thereafter, we concatenate the identified traces of breathing rates in adjacent time windows to further identify the arriving and leaving time of human subjects.

In Stage 3, we leverage the estimated breathing rate traces given by Stage 2 to do people counting and recognition. By further utilizing the time-domain information and removing outliers of the estimated breathing rate traces by a novel quasi-bilateral filter, the system gives an estimate of the crowd number.

3.3 Multi-user Breathing Spectrum Generation

The breathing signals can be extracted by applying *Short-Term Fourier Transform* (STFT) to the CSI measurement. In specific, we first apply a sliding window of length W to the captured CSI time series of each SC in every link, and then obtain the frequency spectrum by performing *Fast Fourier Transform* (FFT) over each time window. We then employ a band-pass filter on the spectrum to consider only the normal range of human breathing frequencies $[b_{\min}, b_{\max}]$. The FFT is performed on every SC to obtain the individual spectrum for all the $N_{\text{Tx}} \times N_{\text{Rx}} \times N_{\text{sc}}$ SCs, where N_{Tx} , N_{Rx} , and N_{sc} are the number of Tx antennas, Rx antennas, and usable SCs on each Tx-Rx link, respectively.

As shown in Fig. 3.2, each breathing signal from one person contributes to one evident peak in the obtained Power Spectrum Density (PSD). Note that different SCs experience diverse sensitivity levels to the identical breathing motion. Previous approaches attempt to select a set of best SCs based on variance, amplitude or ensemble average of CSI among all SCs to improve SNR. However, the following observations show the flaws of these approaches: 1) The response power of different SCs to the same breathing source is different (See columns in Fig. 3.2). 2) For the same SC, the response power to different breathing sources is different (See rows in Fig. 3.2). 3) The response power of different links is distinct (Different figures in Fig. 3.2). Therefore, there is no single SC that is universally sensitive to all breathing sources. Using the same subset of SCs for different frequency components may not produce equally good SNR for all breathing signals.

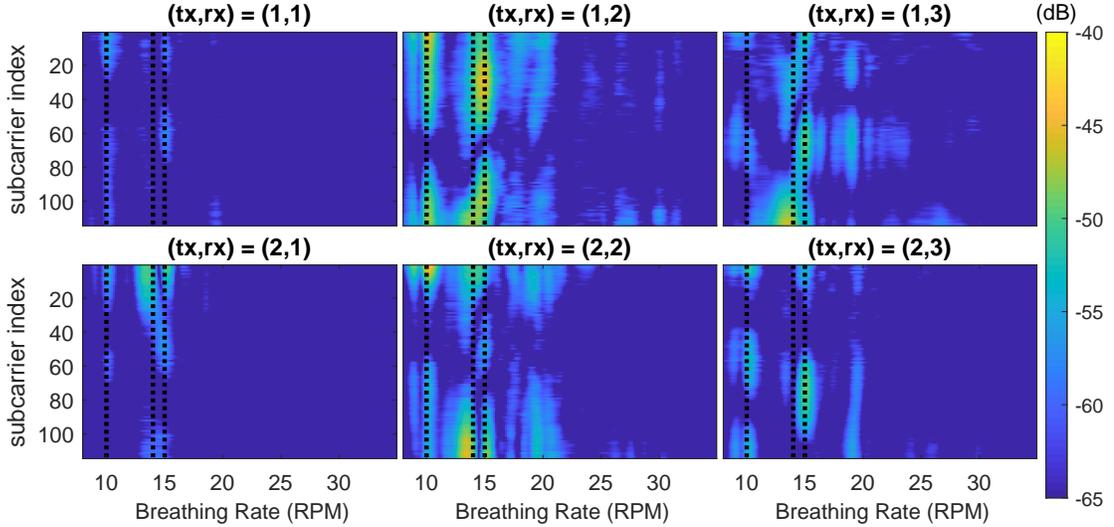


Figure 3.2: PSD of different links. Ground-truth: 3 people sitting in car with breathing rate [10 14 15] Respiration Per Minute (RPM) (marked as dashed lines). Furthermore, using a universal threshold for all links may lose information from links with low response power.

Inspired by these observations, we first propose a novel adaptive SC combining criteria to boost the SNR of breathing signal of each link. For link m , the selected SCs for a given frequency component q satisfy the condition that

$$E_k^{(m)}(q) \geq \alpha \max_{q \in Q, i \in \mathcal{V}} \{E^{(m)}(q, i)\}, \forall k \in \mathcal{V}, \quad (3.1)$$

where Q is the set of frequency components in the range of $[b_{\min}, b_{\max}]$. $E_k^{(m)}(q)$ denotes the power of the k -th SC over link m at frequency component q and $\max_{q \in Q, i \in \mathcal{V}} \{E^{(m)}(q, i)\}$ denotes the maximum power of link m over all frequency components and SCs. α is a hyper-parameter which determines a relative threshold $th^{(m)} = \alpha \max_{q \in Q, i \in \mathcal{V}} \{E^{(m)}(q, i)\}$ for SC selection. Note that $th^{(m)}$ is adaptive

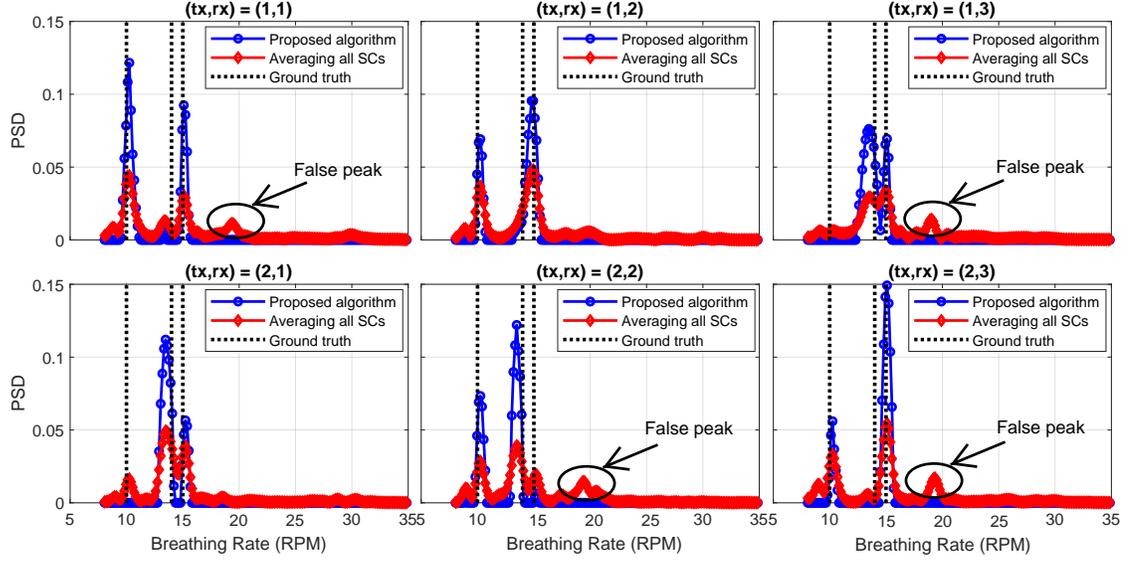


Figure 3.3: Normalized PSD of different links after SC combination.

to individual link quality, as inspired by the third observation above. Thus, the enhanced power of frequency component q in link m is

$$E^{(m)}(q) \leftarrow \sum_{k \in \mathcal{V}} E_k^{(m)}(q) \mathbb{1}(E_k^{(m)}(q) \geq th^{(m)}). \quad (3.2)$$

To further incorporate diverse link quality, we normalize the power for each link and then combine them together to further improve the SNR:

$$E^{(m)}(q) \leftarrow \frac{E^{(m)}(q)}{\sum_{i \in Q} E^{(m)}(i)}, \forall q \in Q, \quad (3.3)$$

$$E(q) \leftarrow \sum_{m=1}^M E^{(m)}(q), \forall q \in Q, \quad (3.4)$$

where $E(q)$ is the power of frequency component q after link combination and $M = N_{\text{Tx}} \times N_{\text{Rx}}$ is the total number of links.

Fig. 3.3 shows the effect of SC combination for several exemplary links. As seen, the proposed SC selection and combination scheme (shown in blue curves)

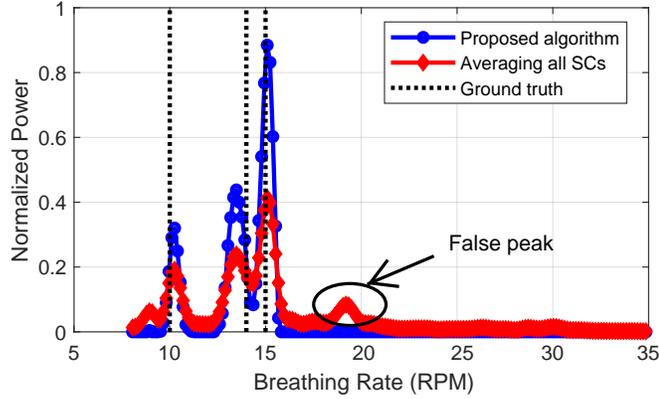


Figure 3.4: PSD after link combination.

remarkably improves the SNR for the frequency components of interests, outperforming the simple average scheme (shown in red curves). Fig. 3.4 further depicts the PSD after the combination of all 9 links, which demonstrates that noise and interference have been effectively suppressed. The ground-truth of breathing rates are marked with the black dashed lines. As a comparison, simple average of all SCs suffer from less dominant peaks for the desired breathing signals and false peaks.

3.4 Breathing Rate Trace Tracking

3.4.1 From Breathing Rates to Breathing Rate Traces

Previous works estimate the number of people by the number of candidate breathing rates [11]. However, they have several limitations. First, the breathing rate estimation may not be accurate enough for a single time instance. Second, different users may have close breathing rates that are indistinguishable from the

frequency spectrum, resulting in potential underestimation. Third, the number of people could vary over time as people may come and go. And the accompanying motion will also corrupt the breathing signals.

To map the breathing rates to individuals and thus further estimate the accurate crowd number, we utilize the diversity in the time series of breathing rate estimates for reliable estimation. We first model the dynamic breathing rates as a Markov process. Noting that the breathing signals are periodic where breathing frequency can smoothly change over time, the variation of breathing rate between two adjacent time bins is assumed to follow a normal distribution $\mathcal{N}(0, \sigma^2)$, with the Probability Density Function (PDF) $p(f)$. Since the operation of STFT automatically discretizes the continuous frequency in the range of $[b_{\min}, b_{\max}]$ into $|Q|$ frequency components, where $|Q|$ means the cardinality of set Q , the natural breath can be modeled as a Markov chain, and the transition probability matrix is denoted as $\mathbf{P} \in \mathbb{R}^{|Q|} \times \mathbb{R}^{|Q|}$, which is defined as

$$\begin{aligned} \mathbf{P}(q, q') &= \mathbf{P}(g(i) = q' | g(i-1) = q) \\ &= \int_{(q' - q - \frac{1}{2}) * \Delta f}^{(q' - q + \frac{1}{2}) * \Delta f} p(f) \mathbf{d}f, \end{aligned} \tag{3.5}$$

where $\forall q, q' \in Q$ and g is a mapping indicating the frequency component of the breathing rate at given time slots.

To estimate the breathing rate trace in a given time slot t , our system leverages the spectrum in $[t - W, t]$, where W is the window length. An output is produced every W_s seconds, and the spectrum is updated at the same time. Thus to estimate the breathing traces at time t , a spectrum $S \in \mathbb{R}_+^I \times \mathbb{R}_+^{|Q|}$ is leveraged,

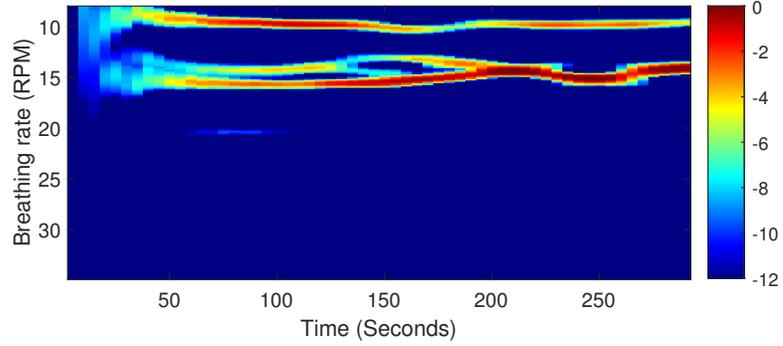


Figure 3.5: Spectrogram after link combination.

where $I = \frac{W}{W_s}$.

In principle, the breathing signal is more periodic than noise and other motion interference. Thus, it is more likely to be observed as peaks in most of the time, and thus the breathing signal will form a trace in the given spectrum along the time with the frequency changing slightly, as shown in Fig. 3.5. In the following, we first extract the traces of successive breathing rates in the given window, and then concatenate them over time.

3.4.2 Extracting Breathing Rate Traces

3.4.2.1 Theoretical Model

For a given spectrum S , a reasonable estimate of the breathing trace can be obtained by

$$\mathbf{g}^* = \arg \max_{\mathbf{g}} E(\mathbf{g}), \quad (3.6)$$

where \mathbf{g} indicates the breathing trace, denoted as

$$\mathbf{g} = (g(n), n)_{n=1}^I. \quad (3.7)$$

Here, $g : [1, I] \rightarrow Q$ is a mapping indicating the frequency component of the trace at the given time. $E(\mathbf{g})$ is the power of a trace, defined as

$$E(\mathbf{g}) = \sum_{i=1}^I S(i, g(i)), \quad (3.8)$$

where $S(i, j)$ denotes the power at time bin i and frequency component j .

Considering that one's breathing rate will not fluctuate a lot within a short period, a regularization term is added to penalize sudden changes in frequencies of interests. A breathing trace is then a series of breathing rate estimates that achieve a good balance between frequency power and temporal smoothness. The smoothness of a trace can be evaluated by a cost function $C(\mathbf{g})$, defined as

$$C(\mathbf{g}) \triangleq -\log \mathbf{P}(g(1)) - \sum_{i=2}^I \log \mathbf{P}(g(i-1), g(i)), \quad (3.9)$$

where the frequency transition probability $\mathbf{P}(g(i-1), g(i))$ can be calculated by (3.5). Without loss of generality, we assume a uniform prior distribution, i.e., $\mathbf{P}(g(1)) = \frac{1}{|Q|}$. The cost function $C(\mathbf{g})$ is the negative of the log-likelihood for a given trace. The smoother a trace is, the larger its probability is, and the smaller the cost it incurs.

The most probable breathing trace can be found by solving

$$\mathbf{g}^* = \arg \max_{\mathbf{g}} E(\mathbf{g}) - \lambda C(\mathbf{g}), \quad (3.10)$$

where λ is a regularization factor. Here we denote $E(\mathbf{g}) - \lambda C(\mathbf{g})$ as the regularized energy of trace \mathbf{g} . By properly choosing the hyper-parameter λ , the system can ensure that the regularized energy of a true breathing trace is positive, while

when the observation area is empty, the regularized energy for any trace candidate in the given spectrum is negative.

3.4.2.2 Iterative Dynamic Programming

The problem in (3.10) can be solved by dynamic programming. However, dynamic programming typically can only find the trace with the maximum regularized energy and cannot deal with multiple breathing traces. We propose a successive cancellation scheme to find multiple traces one by one via a novel method of iterative dynamic programming (IDP).

The principle idea of the IDP is intuitive. For a given spectrum S , the most probable breathing trace is first found by dynamic programming. To further determine if there are any other breathing traces, the identified trace will be erased from the spectrum, and then a new round of dynamic programming is performed to find another candidate trace. This successive cancellation procedure will be run iteratively until there is no more effective traces in the spectrum.

For clarity, (i, q) denotes the bin index with timestamp i and frequency component q . We want to find the best trace of frequency peaks from timestamp i to j , which is denoted as $\mathbf{g}_i \rightsquigarrow \mathbf{g}_j$. Define the regularized energy of trace $\mathbf{g}_i \rightsquigarrow \mathbf{g}_j$ that ends at point (j, n) as $s(\mathbf{g}_i \rightsquigarrow (j, n))$. Our approach is to search all possible traces $\mathbf{g}_i \rightsquigarrow (j, n)$ that end at frequency point n and select the best one among them. This can be achieved by finding the optimal traces for all the bins along with the adjacent timestamps. For simplicity, we denote the regularized energy

at each bin as its score given by

$$s(i, q) = S(i, q) + \max_{\forall q' \in Q} \{s(i-1, q') + \lambda \log \mathbf{P}(q', q)\}, \quad (3.11)$$

$$i = 2, 3, \dots, I, \forall q, q' \in Q,$$

where $s(1, q) = S(1, q) + \lambda \log \mathbf{P}(g(1) = q)$. The score of a given bin is the maximum achievable regularized energy that it can obtain. In other words, it determines the optimal paths that pass through bin (i, q) .

Algorithm 1 Iterative Dynamic Programming

- 1: Calculate regularized energy map $s(i, j)$ by (3.11)
 - 2: Initialize trace number $t \leftarrow 0$, frequency response of rectangular window \mathbf{w}
 - 3: **while** $\max_q s(I, q) > 0$ **do**
 - 4: $t \leftarrow t + 1$
 - 5: $g_t(I) \leftarrow \arg \max_q g(I, q)$
 - 6: $i \leftarrow I - 1$
 - 7: **while** $i \neq 0$ **do**
 - 8: $g_t^*(i) \leftarrow \arg \max_q s(i, q) + \lambda \mathbf{P}(q, g_t^*(i+1))$
 - 9: $i \leftarrow i - 1$
 - 10: **end while**
 - 11: update spectrum $\mathbf{S}(i) = \mathbf{S}(i) - S(i, g_t^*(i)) * \mathbf{w}, \forall i = 1, 2, \dots, I$
 - 12: Calculate regularized energy map $s(i, j)$ by (3.11)
 - 13: **end while**
-

The entire optimal breathing trace can be found by backtracking the bins that contribute to the maximum score $g^*(I)$ of the last timestamp. For the rest of

the breathing trace in the observation window, i.e., $\forall i = I - 1, I - 2, \dots, 1$, we have

$$g^*(i) = \arg \max_{\forall q \in Q} s(i, q) + \lambda \log \mathbf{P}(q, g^*(i + 1)). \quad (3.12)$$

The backtracking procedure in (3.12) gives the optimal trace \mathbf{g}^* for a given spectrum, which is the optimal solution for (3.10).

To further check if there are any other candidate breathing signals in the given spectrum, the trace \mathbf{g}^* should be removed. For the ideal case, we only need to remove the bins along \mathbf{g}^* . However, since the number of FFT points are limited, the energy of the breathing signal is diffused around the center of breathing trace, which forms an energy strip in the given spectrum as shown in Fig. 3.5. Thus, if we only remove the energy along the optimal trace \mathbf{g}^* and consecutively execute dynamic programming in (3.11) and (3.12), we will get a group of traces inside one energy stripe. Therefore, IDP applies a windowing module on the optimal trace \mathbf{g}^* to emulate the diffusing effect of FFT to get an energy stripe. The updated spectrum after we erase the optimal energy stripe is

$$\mathbf{S}(i) \leftarrow \mathbf{S}(i) - S(i, g^*(i)) * \mathbf{w}, \forall i = 1, 2, \dots, I, \quad (3.13)$$

where $\mathbf{S}(i)$ denotes the energy of spectrum at timestamp i , and \mathbf{w} is the frequency response of the windowing module. Operator $*$ denotes convolution operation, which can emulate the energy stripe caused by the diffusing effect of FFT.

We recursively perform the above dynamic programming and spectrum cancellation to find multiple traces. The algorithm terminates when the score of the found trace is negative, indicating an empty spectrum without any effective traces.

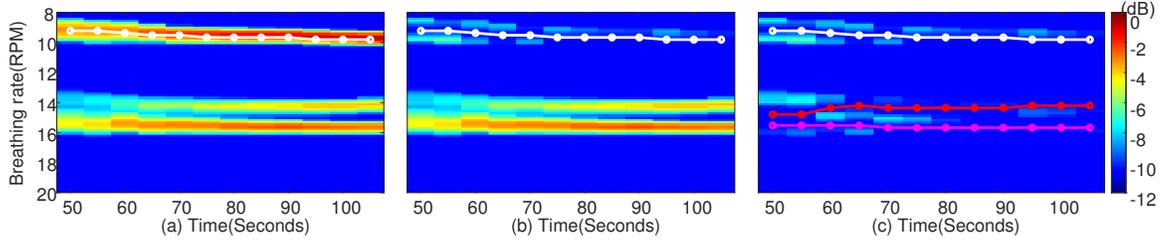


Figure 3.6: Successive cancellation procedure of IDP.

The procedure of iterative dynamic programming is summarized in Algorithm 1. Fig. 3.6 illustrates the details of this finding-then-erasing procedure. In Fig. 3.6 (a), the trace found by DP is marked by the line, and the energy stripe of this trace is removed as shown in Fig. 3.6 (b). The spectrogram, when IDP terminates, is shown in Fig. 3.6 (c), and lines in the figure indicate the breathing traces. It is clear to see that although there is still some residual energy not perfectly removed, IDP terminates properly since there are no traces satisfying the constraint of non-negative regularized energy.

3.4.2.3 Detecting Empty Case

Ideally, when there is no person present in the monitoring area, no breathing trace would be picked up since the spectrum would be random due to the normal distribution of the thermal noise. In reality, however, false traces could be detected since some noise might be boosted in the empty case. To avoid this effect, we employ motion detection to determine empty cases. If no motion (not even chest movement) is detected, the system will directly claim empty; otherwise, the above steps are performed to find a potential breathing rate trace. Here

the motion detector needs to be sensitive and robust enough to detect breathing motion. In this paper, we employ the state-of-the-art approach proposed in [90] for this purpose, which achieves almost zero false alarm.

3.4.3 Trace Concatenating

Iterative dynamic programming provides the breathing rate traces for each time window. In practice, a continuous monitoring system, however, would operate for much longer time than a time window, posing extra information gains to enhance the trace extraction. In this part, we propose a novel trace concatenating algorithm to concatenate trace segments belonging to the same breathing signal in different time windows, which not only improves the trace segments, but also enables detection of the start and end time of each trace (or equivalently, the entering and leaving time of a specific user).

For clarity, we store all presented traces in a database. The j th trace found previously is denoted as $\mathbf{g}_j^{\text{pre}}(t_{\text{st}} : t_{\text{end}})$, where $j = 1, \dots, J$ and t_{st} and t_{end} denote the start and end time of the trace. The k th traces found in the current time window $[t - W, t]$ is denoted as $\mathbf{g}_k(t - W : t)$, where $k = 1, \dots, K$. Furthermore, the similarity between two traces is defined as the ratio between the overlapped time in the window and the window length, which is expressed as

$$f(\mathbf{g}_j^{\text{pre}}, \mathbf{g}_k) = \frac{|\mathbb{1}(\mathbf{g}_j^{\text{pre}}(t_{\text{st}} : t_{\text{end}}) = \mathbf{g}_k(t - W : t))|}{I - 1}, \quad (3.14)$$

where $f(\mathbf{g}_j^{\text{pre}}, \mathbf{g}_k) \in [0, 1]$. A similarity matrix $F \in \mathbb{R}^J \times \mathbb{R}^K$ can be calculated accord-

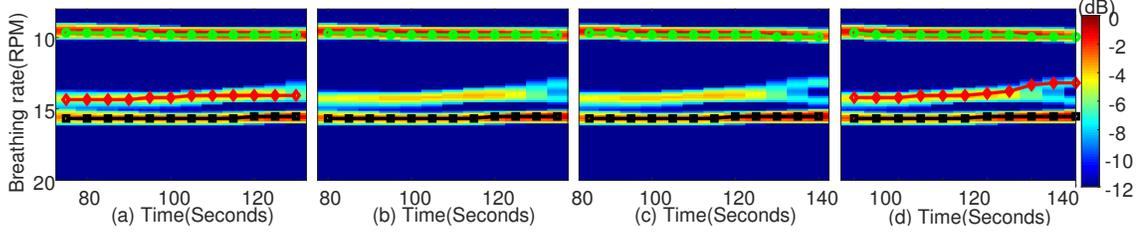


Figure 3.7: Traces found by IDP in four adjacent time windows.

ing to (3.14) to show the similarity between all the traces in the current window and those in the database. In order to find the previous part for $\mathbf{g}_k(t - W : t)$, we only need to find the maximum item of $\mathbf{f}(\mathbf{k})$, which is the k -th column of F . The row index of the maximum similarity indicates the index of the previous trace if the maximum similarity is above a predefined threshold. If there exists a previous trace with a high enough similarity, it means that the corresponding breathing rate trace has been detected before. Then the endpoint of the corresponding trace should be updated. We let the endpoint be the current time and update the corresponding frequency component accordingly. In case a new user arrives, there will be no existing traces that have a similarity larger than the threshold and thus a new trace is created with the corresponding timestamps and frequency components. Similarly, no trace in the current window being similar to the past traces corresponds to a user that has left, and thus the trace would be terminated.

Fig. 3.7 and Fig. 3.8 show the effect of trace concatenating algorithm. Four adjacent time windows are shown in Fig. 3.7, and traces found by IDP are marked by lines. We can see that although the breathing trace in the middle of the spec-

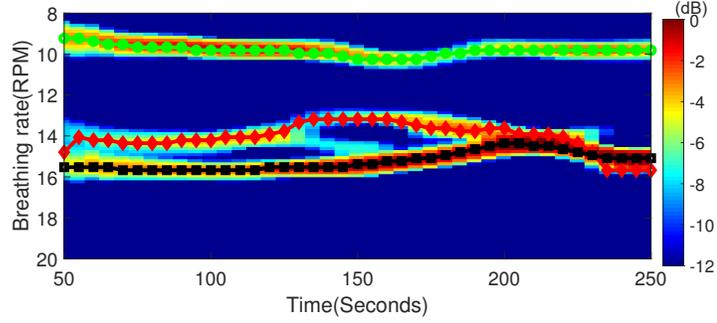


Figure 3.8: Trace concatenating result of windows in Fig. 3.7.

rogram is not detected in the second and third window (due to body motion as well as breathing rate change of the subject), since the trace found in the fourth window still achieves high similarity with the trace found in the first window, it still can be concatenated as shown in Fig. 3.8.

3.5 People Counting

IDP and trace concatenation provides the estimation of breathing rate traces, and the trace number would be the estimate of occupancy level. Although the IDP and trace concatenating have considered the stability of human breath in the observation area, the estimation result may still suffer from large noise and have some false alarms or underestimations for a real-time output as shown in Fig. 3.7 (b) (c). To eliminate/mitigate these outliers for a real-time system, we design a quasi-bilateral filter to explore the information contained in the previous estimations. Similar to the bilateral filter [71], the designed filter considers the distance in time and estimation domain simultaneously, but we make some improvements according to our system. First, for a real-time system, the filter

can only use the past and current estimation result. Furthermore, since IDP and trace concatenation leverage time as well as frequency information to get continuous breathing rate traces, the preliminary estimations are consistent in a short period. Thus, if we directly use a bilateral filter, only the first incorrect output will be rectified. Given these two constraints, we develop a segment-based filter, where each segment is a consistent preliminary estimation sequence.

Specifically, the output is determined by the current estimation and the previous segments. We denote the weight of segment s as $W_{\text{seg}}(s)$ expressed as

$$W_{\text{seg}}(s) = w(l_s) * w(\tau_s) * w(d_s), \quad (3.15)$$

where l_s is the length of segment s , and τ_s is the time difference between segment s and current timestamp. d_s is the estimation difference between current estimation and segment s as shown in Fig. 3.9. Intuitively, the longer the segment is, the greater weight will be assigned. In contrary, the larger the distance is, no matter in time or the estimate of the crowd number, the smaller the influence of this segment imposing on the current result. For clarity, the set of segments with i estimated people is denoted as S_i , and the current estimated number as j . The weight that the currently estimated people is i after quasi-bilateral filter can be calculated by

$$p(i) \leftarrow \frac{1}{N} \left(\sum_{s \in S_i} \frac{l_s}{\tau_s} \right) e^{-d_s}, \quad (3.16)$$

where N is the total number of segments, the estimation difference is $d_s = |i - j|$, and $W_{\text{seg}}(s)$ in (3.15) is designed as

$$W_{\text{seg}}(s) \leftarrow \frac{l_s}{\tau_s} e^{-d_s}. \quad (3.17)$$

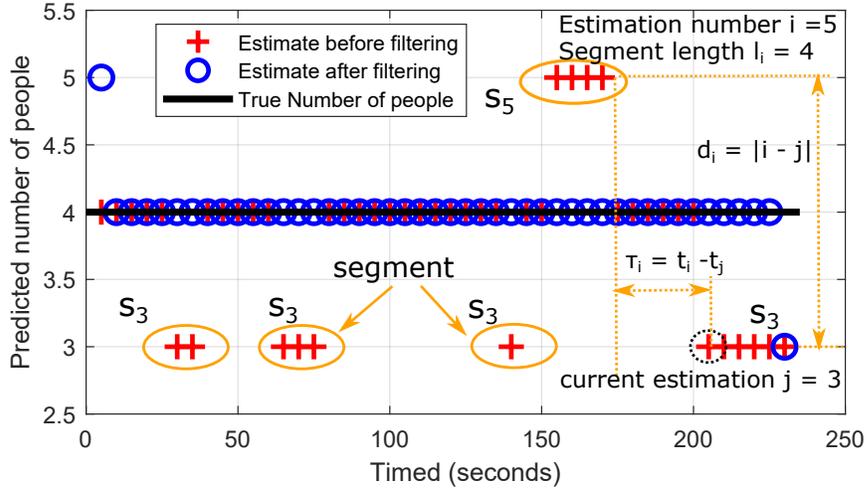


Figure 3.9: Example and illustration of quasi-bilateral filter.

The eventual result after filtering is j' , given by

$$j' = \arg \max_i p(i). \quad (3.18)$$

Fig. 3.9 shows the estimation results before and after quasi-bilateral filtering. Clearly, the novel quasi-bilateral filter can remove the estimation outliers effectively, and thus improve the performance of people counting system.

3.6 Experiments and Evaluation

In this section, we conduct extensive experiments to evaluate the performance of the proposed approach. Specifically, we first introduce the experimental setup and then the results corresponding to two different scenarios. Discussion on the impact of distinct modules proposed in Section 3.4 follows.

3.6.1 Device and Methodology

We conduct experiments using a pair of commodity WiFi devices, one as Tx and the other as Rx. The channel is set to 5.765 GHz with a bandwidth of 40 MHz. Both Tx and Rx are equipped with 3 omnidirectional antennas. Each link between a Tx antenna and an Rx antenna has a total of 114 SCs. Considering for practical long-term monitoring, we use a very low sampling rate of 10 Hz.

All the data in our experiments are collected in an on-campus lab and a car over two months with 12 participants. Fig. 3.10 (a) shows the layout of the LAB in which two devices (Tx and Rx) are put on two different sides of a round desk, and the distance between the Tx and Rx is 3.5 m. Participants are invited to sit in chairs as if they were attending a meeting. During the experiments, the participants randomly choose their seats and slight movements are allowed. To further verify that the proposed system is independent of the environment, we also conduct experiments in a car, which is an extreme case for indoor scenario, where there is limited space as well as strong reflection. For the car scenario, the Tx and Rx are put at the front door on the driver and passenger side respectively, as shown in Fig. 3.10 (b).



(a) LAB

(b) Car

Figure 3.10: Experiment setup.

3.6.2 Overall Performance

Fig. 3.11 (a) shows the confusion matrix of our method in the LAB, and the overall accuracy is 87.14%, with the accuracy defined as

$$\text{Accuracy} = \frac{\# \text{ of predicted label equals true label}}{\text{total \# of samples}}. \quad (3.19)$$

The counting error is within 1 person for 98.6% of the testing cases. Additionally, the proposed system can perfectly detect whether the monitoring area is occupied or not. The accuracy however, decreases with more people present. This is as expected since the more people there are, the more likely their breathing traces may merge together and the more likely occasional motion may happen, both leading to counting errors. Fig. 3.12 (a) shows that our testing result in the car can achieve a comparable performance with that in the LAB, which demonstrates the independence of our system on the environment.

To further evaluate our system, we compare it with the most relevant TR-BREATH [11] which also estimates multi-person breathing rates using commer-

Predicted Number	0	1	2	3	4
0	100.0%	0.0%	0.0%	0.0%	0.0%
1	0.0%	100.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	94.7%	24.2%	1.4%
3	0.0%	0.0%	5.3%	75.8%	17.4%
4	0.0%	0.0%	0.0%	0.0%	65.2%
5	0.0%	0.0%	0.0%	0.0%	15.9%
	0	1	2	3	4
	True Number				

(a) Proposed system

Predicted number	0	1	2	3	4
0	95.8%	0.0%	14.8%	0.0%	0.0%
1	0.0%	89.7%	14.8%	11.4%	0.0%
2	4.2%	10.3%	66.7%	12.7%	3.9%
3	0.0%	0.0%	3.7%	55.7%	15.6%
4	0.0%	0.0%	0.0%	12.7%	45.5%
5	0.0%	0.0%	0.0%	7.6%	22.1%
6	0.0%	0.0%	0.0%	0.0%	9.1%
7	0.0%	0.0%	0.0%	0.0%	2.6%
8	0.0%	0.0%	0.0%	0.0%	1.3%
	0	1	2	3	4
	True number				

(b) TR-BREATH system

Predicted Number	0	1	2	3	4
0	100.0%	0.0%	0.0%	0.0%	0.0%
1	0.0%	100.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	95.7%	24.2%	2.2%
3	0.0%	0.0%	4.3%	73.2%	21.7%
4	0.0%	0.0%	0.0%	2.6%	58.7%
5	0.0%	0.0%	0.0%	0.0%	17.4%
	0	1	2	3	4
	True Number				

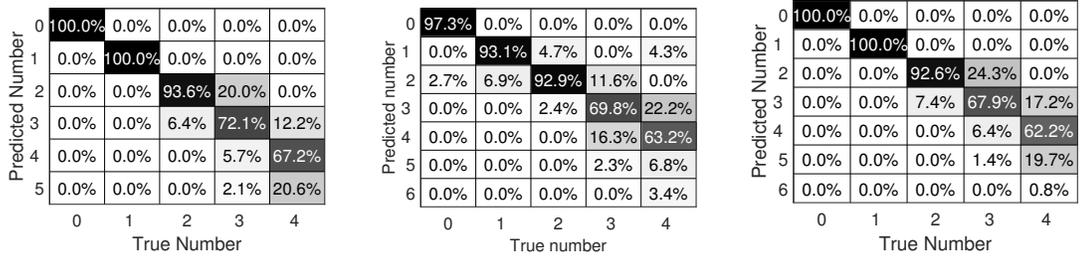
(c) Without quasi-bilateral filter

Figure 3.11: Confusion matrix of people counting in LAB.

cial WiFi. TR-BREATH employs root-MUSIC for breathing rate estimation and uses the affinity propagation algorithm to estimate crowd number. In order to make fair comparison, quasi-bilateral filter is used to filter out the outliers of original TR-BREATH estimations. The estimation accuracy of TR-BREATH [11] in LAB and car are shown in Fig. 3.11 (b) and Fig. 3.12 (b) respectively. As seen, TR-BREATH shows a comparable performance in the car testing scenarios. The performance in the LAB environments is much worse, with an overall accuracy of 70.68%. The proposed approach improves the overall performance by 16.46% and 3.32% for LAB and car testing scenario respectively, attributed to its three core techniques: adaptive SC combination, iterative dynamic programming, and trace concatenation.

3.6.3 Performance Gain of Individual Modules

In this section, we discuss how each independent module improves the performance of our system. Apart from the confusion matrix and accuracy, here we



(a) Proposed system (b) TR-BREATH system (c) Without quasi-bilateral filter

Figure 3.12: Confusion matrix of people counting in car.

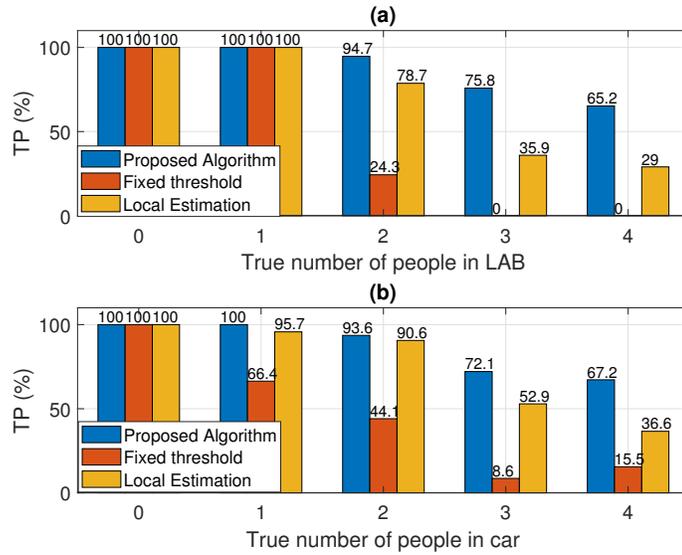


Figure 3.13: TP comparison of different algorithms.

additionally adopt true positive (TP) rate, which is calculated as:

$$TP_i = \frac{\# \text{ of samples that predicted label is } i}{\text{total } \# \text{ of samples that true label is } i}. \quad (3.20)$$

3.6.3.1 Impact of SNR Boosting Algorithm

Here, we compare the proposed SNR boosting algorithm with the commonly used one, i.e., selecting the SCs whose maximum energy are above a cer-

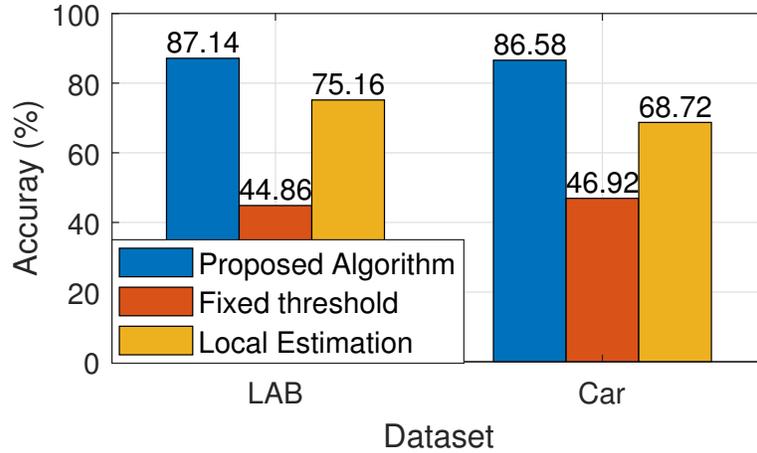


Figure 3.14: Accuracy of different algorithms.

tain threshold (hereafter, it is called *fixed threshold* algorithm). For fair comparison, we choose the 30% of the maximum link energy as the threshold for both of the methods. Furthermore, the energy of each link is normalized before link combination, thus the parameters used in later process for both methods are also the same. Fig. 3.13 compares the TP of the two SC selection algorithms. It can be easily seen that our proposed algorithm shows better performance. This is benefited from our observation that SC has different sensitivity on distinct breathing signals, i.e., for each SC, we only choose the part that is most likely to be a signal, thus compressing noise and interference well. Fig. 3.14 also shows the superiority of our method in the accuracy aspect. Besides, the fixed threshold algorithm degrades significantly when there are more than 3 people because they cannot effectively suppress noise and interference.

3.6.3.2 Impact of IDP Estimation Algorithm

In this experiment, we show the benefits of the proposed IDP. We compare the performance with a *local estimation* algorithm that estimates the number of people based on the spectrum at current timestamp only.

For fair comparison, the quasi-bilateral filter is also applied to the local estimation algorithm. The comparisons of TP and accuracy for the two algorithms are portrayed in Fig. 3.13 and Fig. 3.14, respectively. The results shows that IDP considerably improves the performance for both datasets, which demonstrates the gains contributed by leveraging time diversity in counting.

3.6.3.3 Impact of Quasi-bilateral Filter

In this experiment, we show the effect of the designed quasi-bilateral filter on the performance of the people counting system. Fig. 3.11 (c) and Fig. 3.12 (c) shows the confusion matrix of people counting system without filtering on datasets collected in LAB and car respectively. By comparing the result with Fig. 3.11 (a) and Fig. 3.12 (a), we can see that the quasi-bilateral filter can improve the performance in most cases, especially when the number of people is larger than 3 in the observation area. The reason is that when the number of subjects increases, more motion interference will be introduced. Furthermore, it is more likely that different breathing traces will merge. Even though we utilize the time domain as well as frequency diversity by IDP, estimation error still can occur. Quasi-bilateral filter is a post-processing method that will further utilize the diversity in

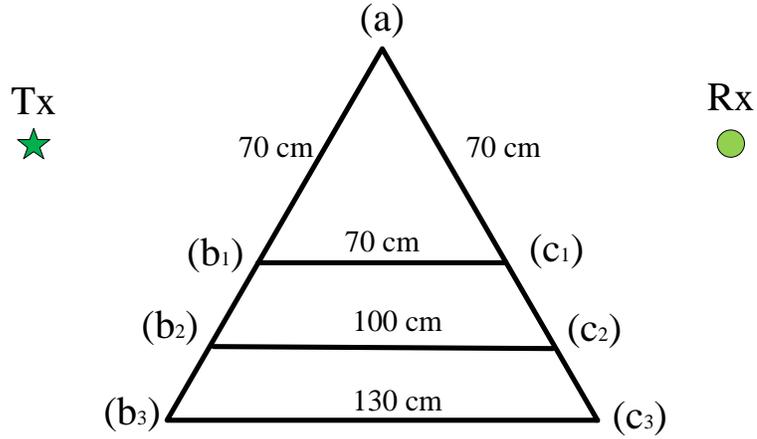


Figure 3.15: Experiment setup for resolution investigation.

time domain and thus correct the estimate outliers.

3.6.4 Resolution Discussion

To further investigate the impact of spatial separation as well as respiration rate difference of human subjects, we perform experiments with 3 participants sitting with different spatial separations, as shown in Fig. 3.15. Considering the volume of a human subject, the minimum distance is set as 70 cm. The distance between Tx and Rx is 3.2 m. To ensure a constant breathing rate separation during the experiments, each of the subjects performs controlled breathing according to a metronome. The breathing rate separations of [0.5, 1, 1.5, 2] RPM are evaluated, respectively. Table. 3.1 and Table. 3.2 show the performance of our system, where the 4-tuple (*;*;*;*) denotes the detection accuracy and the relative breathing rate accuracy with the 3 users at location a , b , and c respectively. We can see that the breathing rate separation has a significant impact on the per-

Frequency separation		Spatial separation	
		0.5 RPM	1 RPM
70 cm		(65.5;98.7;98.0;98.7) ¹	(96.5;98.9;96.5;97.3)
100 cm		(62.1;98.6;99.4;99.2)	(96.6;98.6;98.4;98.9)
130 cm		(65.0;98.1;98.5;99.7)	(96.6;98.9;98.5;97.7)

¹ The 4-tuple (*;*;*;*) denotes the detection accuracy and the relative breathing rate accuracy with 3 participants at location *a*, *b*, and *c* respectively. For example, (65.5;98.7;98.0;98.7) denotes the detection accuracy is 65.5%, and the breathing rate accuracy at location *a*, *b*, and *c* are 98.7%, 98.0%, and 98.7% respectively.

Table 3.1: Accuracy with different frequency and spatial separation - Part I.

formance, while the impact of the spatial separation is negligible. The detection rate raises more than 30% when the breathing rates separation increases from 0.5 RPM to 1 RPM. The system achieves 100% detection rate once the breathing rate separation is above 1.5 RPM. Besides, as long as the breathing rate trace has been detected, our system can accurately estimate the breathing rate, and the breathing estimation accuracy is above 94.7% for all of the test case.

3.7 Summary

This chapter presents a breathing rate trace tracking system using commercial WiFi. The proposed system enables static crowd counting by multi-person

Frequency separation		Spatial separation	
		1.5 RPM	2 RPM
70 cm	(100;97.3;98.8;97.7)	(100;98.1;98.9;94.7)	
100 cm	(100;97.8;99.0;98.9)	(100;98.2;98.5;99.0)	
130 cm	(100;98.4;99.1;97.1)	(100;98.1;99.3;99.4)	

Table 3.2: Accuracy with different frequency and spatial separation - Part II.

breathing rate trace tracking with three key components: an adaptive subcarrier combination method that boosts breathing signals, an iterative dynamic programming algorithm to extract the successive breathing traces from different individuals, and a trace concatenating algorithm that splices consecutive breathing trace segments. Experiment results show a respective average accuracy of 87.14% and 86.58% for an indoor office and car environments for people counting. Additionally, the counting error is within 1 person for 97.9% of the time.

Chapter 4: Multi-person Respiration Rate and Heart Rate Monitoring

4.1 Introduction

Over the past decade, great efforts have been put into designing and testing different architectures for robust vital sign monitoring using off-the-shelf WiFi devices [11, 12, 34, 82, 87, 88, 91]. However, due to the relative low carrier frequency of WiFi systems, the antenna number of 2.4/5GHz WiFi is small, rendering a low spatial resolution. Besides, the narrow bandwidth of WiFi systems results in a coarse range resolution (7.5 meters with bandwidth 20MHz). Therefore, when there is more than one user present, the received RF signals are reflected by the multiple users and it is hard to extract the vital signs for each of them. Thus, most of the works assume there is a single person [82, 87, 91], or the breathing rates of different users are distinct [11, 12, 88]. Moreover, since the perturbation caused by the heartbeat is very small (i.e., 0.2 ~ 0.5 mm [57]), the embedded heartbeat signal has an extremely low Signal-to-Noise Ratio (SNR). It is extremely difficult, if possible, to use commodity WiFi to estimate the heart rate [34].

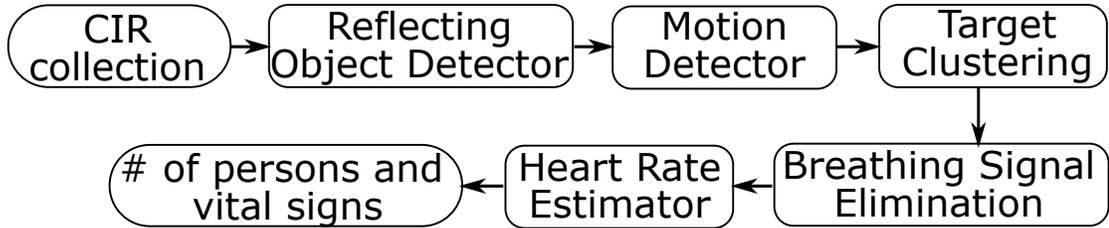


Figure 4.1: An overview of ViMo.

In this chapter, we break down the limitation by leveraging an opportunity in the emerging 60GHz WiFi (e.g., 802.11ad [2]), which is already available in commercial routers [55]. We present **ViMo**, the first system that achieves multi-person stationary/non-stationary detection and vital signs monitoring using an impulse-based commodity 60GHz millimeter wave (mmWave) device. Different from 2.4GHz/5GHz radios, 60GHz WiFi offers high directionality with large phased arrays in small size thanks to millimeter-wavelength and precise time-of-flight measurements brought by the large bandwidth. The advance in 60GHz radio allows higher spatial resolution and range resolution, making it possible to monitor respiration as well as heart rate for multiple persons simultaneously.

The rest of this chapter is organized as follows. We overview the system in Section 4.2, followed by multi-person detection in Section 4.3 and heart rate estimation in Section 4.4. The performance is evaluated in Section 4.5. This chapter is summarized and overviewed in Section 4.6.

4.2 System Overview

ViMo is a wireless system that can accurately detect human subjects and estimate their vital signs by using purely the reflections of RF signals off the users' bodies. The processing flow of ViMo is shown in Fig. 4.1.

Enabling multi-person contactless vital sign monitoring using 60GHz WiFi faces several challenges. First, due to the fast attenuation of 60GHz RF signal [39], the strength of signal reflected at a large distance is much smaller than that at a short distance. Therefore, it is hard to detect human subjects without prior calibration, let alone detecting the stationary/non-stationary status of human subjects. Second, the minute heartbeat signals are easily corrupted by measurement noises and concealed by the large scale respiration signals. Thus, dedicated systems should be designed to resist the interference from respiration and measurement noises when estimating the heart rate.

In order to detect human subjects at various distances, we apply a *reflecting object detector* that adaptively estimates the noise level at various distances and thus detects the presence of reflecting objects. To further differentiate the human subjects from static objects, we design a *motion detector* that identifies static objects, stationary human subjects and human with large body motion. A *target clustering* module is implemented to further identify the number of human subjects and their respective locations. Moreover, to make a robust estimate of the heart rate, we first devise a *breathing signal eliminator* to reduce the interference from the respiration signal after the breathing rate is estimated. The elimina-

tor can remove the harmonics of the breathing signal, as well as deal with the spread of the breathing frequency component when the breathing period slightly changes. To tackle with the random measurement noise, we leverage the stationary property of the heart rate and apply dynamic programming to estimate the heart rate utilizing both the frequency and time diversity.

4.3 Target Detection

Since various indoor objects (e.g., wall, desk, etc.) reflect the EM wave, before starting monitoring vital signs, we first need to detect human subjects in the vicinity of the Tx and the Rx. Note that the human subjects may have body motion and thus will change his/her location in the long run, ViMo divides the duration of measurements into multiple blocks, where each block consists CIR measurements of W seconds. Two adjacent blocks overlap by $W - W_s$ seconds, where W_s is the sliding window length.

4.3.1 Reflecting Objects Detector

Since the RF signal at 60GHz attenuates severely with distance [39], the reflected energy from the same object varies with distance. To locate the human subject, we first need to identify which spatial cell has reflecting objects.

The CIR measurement for the case when there is no reflecting object and the case when there is a static reflecting object at cell (θ, ϕ, l) can be expressed as

$$h_{\theta, \phi, l}^{\text{empty}}(t) = \epsilon(t), \quad (4.1)$$

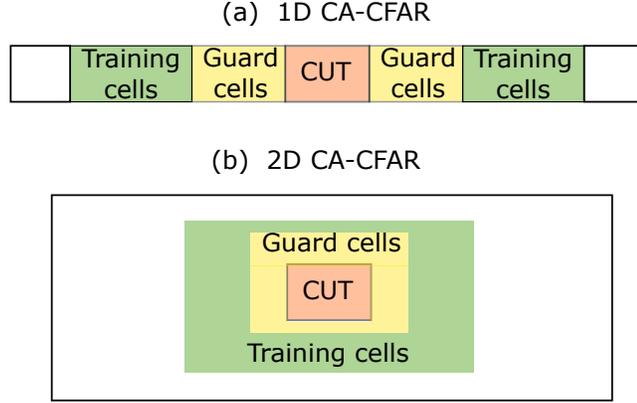


Figure 4.2: CFAR window.

and

$$h_{\theta,\phi,l}^{\text{static}}(t) = \mathbf{s}^H(\theta, \phi) [\mathbf{a} \odot \exp(-j2\pi \frac{\mathbf{d}_0}{\lambda_c})] + \epsilon(t), \quad (4.2)$$

respectively. It is obvious that the power response when there is a reflecting object is much larger than the empty cell. However, it is impossible to find a universal predefined threshold for target detection. According to the propagation laws of EM wave, for the same reflecting object, a shorter distance corresponds to a larger reflecting energy. Furthermore, due to the Automatic Gain Control (AGC) module, the amplitude of the CIRs will change for different measurements.

In order to find the adaptive power threshold for each block, ViMo utilizes Constant False Alarm Rate (CFAR) algorithm [60] for target detection. The power of the noise level for the Cell Under Test (CUT) is estimated by averaging the power of neighboring cells. Furthermore, the guard cell is used to avoid corrupting estimates with power from the CUT itself.

In specific, for each block, the input of CFAR detector is the time-averaging amplitude of all the CIR measurements, i.e., $h(\theta, \phi, l) = \frac{1}{WF_s} \sum_t |h_{\theta,\phi,l}(t)|$, where F_s

is the sampling rate. Considering the attenuation property of EM wave, where the reflected signal strength at a different distance of the same object will be different, to determine the range of the reflecting objects, 1D-CFAR is adopted, as shown in Fig. 4.2 (a). For each sector (θ, ϕ) , ViMo convolves CIR measurements $h_{\theta, \phi}(l)$ with the CFAR window to get the estimation of noise level $\hat{n}_{\theta, \phi}(l)$. A scaling factor α is applied to scale the estimated noise level. The detection threshold is thus set to be $\alpha \hat{n}_{\theta, \phi}(l)$, and the taps with reflecting objects should be those whose amplitude are above the detection threshold, as shown in Fig. 4.3. To determine the noise level at the direction (θ, ϕ) , we further employ 2D-CFAR for tap l , where the noise level $\hat{n}_l(\theta, \phi)$ is estimated by convolving CIR measurements $h_l(\theta, \phi)$ with the 2D-CFAR window as shown in Fig. 4.2 (b). Scaling factor β is applied to scale the estimated noise level. The reflecting object should be in the cell (l, θ, ϕ) whose CIR measurement $h(\theta, \phi, l)$ is above detection threshold $\alpha \hat{n}_{\theta, \phi}(l)$ and $\beta \hat{n}_l(\theta, \phi)$ simultaneously. Here, we define the indicator of reflecting object $\mathbf{1}_R(l, \theta, \phi)$ as

$$\mathbf{1}_R(l, \theta, \phi) = \mathbb{1}\{|h(\theta, \phi, l)| > \max(\alpha \hat{n}_{\theta, \phi}(l), \beta \hat{n}_l(\theta, \phi))\}, \quad (4.3)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Fig. 4.5 shows the cells with reflecting objects in blue circles.

4.3.2 Motion Detector

Although the CFAR detector can identify which cell is occupied by reflecting objects, it cannot differentiate whether the reflection comes from human or not. Note that human subjects always accompany motion (either from breathing

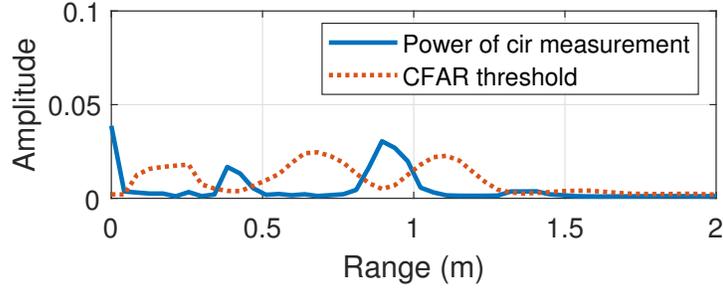


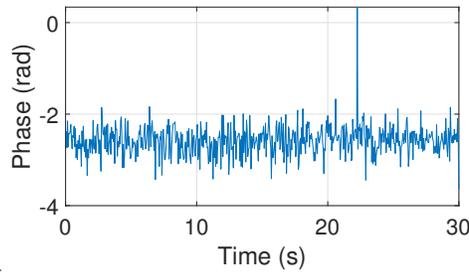
Figure 4.3: Example of 1D-CFAR.

or random body motion), which is a specific characteristic different from static objects, we can design a motion detector to identify human subjects. Furthermore, most of the wireless vital sign monitoring systems assume there is only one human subject and no random body motion during the measurement, and thus the procedure of finding the human subjects is omitted [31, 44, 47, 50, 54, 56, 77], which is neither natural nor realistic for practical deployment. So in this part, we design a motion detector, which enables ViMo to identify static reflecting objects, stationary human subjects and human subject with random body motion.

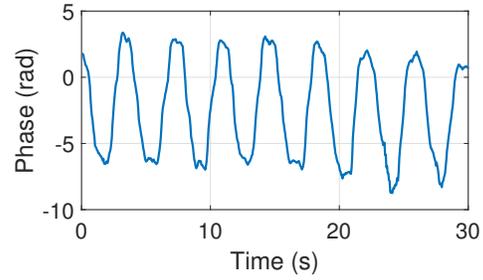
4.3.2.1 Static Reflecting Objects Detection

Note that even stationary human subjects can introduce motion due to respiration and heartbeat, and the distance change caused by respiration can be discerned by phase change according to (2.5), we calculate the variation of the phase of the CIR measurement $V_t(\theta, \phi, l)$ for each candidate cell (θ, ϕ, l) selected by *reflecting objects detector* discussed in Section 4.3.1, which is defined as

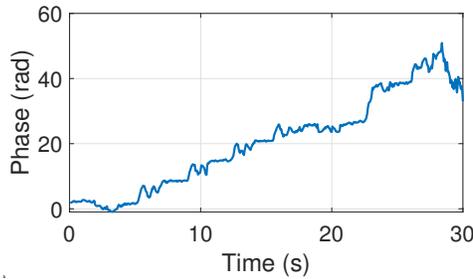
$$V_t(l, \theta, \phi) = \text{Var}_t[\angle h_{\theta, \phi, l}(t)], \quad (4.4)$$



(a) Phase with static object



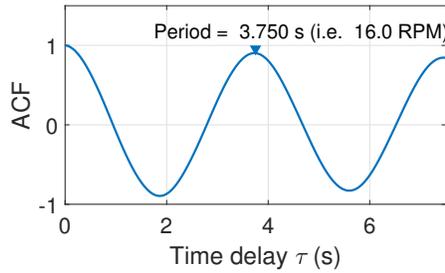
(b) Phase with stationary subject



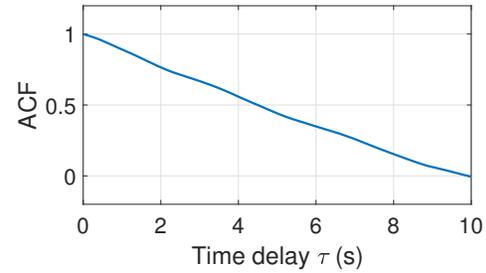
(c) Phase with RBM

State	Phase Variance
Static object	0.099
Stationary subject	12.306
Subject with RBM	199.975

(d) Phase Variance of signals in (a)-(c)



(e) ACF with stationary subject



(f) ACF with RBM

Figure 4.4: Example phase and ACF for target detection.

where $\text{Var}_t[\cdot]$ denotes the variance over parameter t and \angle denotes the phase of a complex value. As shown in (4.2), for a static reflecting objects, $V_t(\theta, \phi, l)$ would be small, but for the cell with human subjects, either respiration or random body motion will contribute a large $V_t(\theta, \phi, l)$. ViMo utilizes a predefined threshold

η_{\min} to identify a static reflecting objects if $V_t(\theta, \phi, l) < \eta_{\min}$ ¹. The phase signal and its variance are shown in Fig. 4.4 (a) and Fig. 4.4 (d) respectively.

4.3.2.2 Stationary Human Subjects Detection

For a stationary human subject, periodic breathing signal can be observed in the phase measurement according to (2.3), and Fig. 4.4 (b) gives an example of a phase measurement with a stationary subject. A bin/cell with a stationary subject would have $V_t(\theta, \phi, l) > \eta_{\min}$ and a periodic phase signal whose frequency within $[b_{\min}, b_{\max}]$.

Note that spectrum analysis is widely used to evaluate the period of respiration in previous works [59]. However, the frequency resolution is $\Delta f = \frac{60}{W}$ Respiration Per Minute (RPM), where W is the window length in seconds. Therefore, to get an acceptable estimation accuracy of the respiration rate, the window length should be long enough, which will cause a large delay. In our system, we adopt a statistical approach by examining the Autocorrelation Function (ACF) of the candidate CIR phase to evaluate the periodicity.

Here we denote the time-variant part of CIR phase measurement as

$$y(t) = s_r(t) + s_h(t) + n(t), \quad (4.5)$$

where $n(t)$ is the random phase offset introduced by noise, and is also a random variable independent in time instances. Thus the ACF of $y(t)$ can be calculated

¹The chest movement caused by tidal breathing ranges from 4-12mm [20]. Considering the wavelength of the 60GHz WiFi, the phase variance is 6.3-56.8. In ViMo, we set $\eta_{\min} = 3$.

as

$$\rho(\tau) = \frac{\text{Cov}[y(t), y(t + \tau)]}{\text{Cov}[y(t), y(t)]}, \quad (4.6)$$

where τ denotes the time lag, and $\text{Cov}[\cdot]$ denotes the con-variance operator. Assume that the distance change caused by heartbeat $s_h(t)$ is uncorrelated with the distance change caused by respiration $s_r(t)$, then $\rho(\tau)$ can be expressed as

$$\rho(\tau) = \frac{\text{Var}[s_r(t)]}{\text{Var}[y(t)]}\rho_r(\tau) + \frac{\text{Var}[s_h(t)]}{\text{Var}[y(t)]}\rho_h(\tau) + \frac{\text{Var}[n(t)]}{\text{Var}[y(t)]}\rho_n(\tau), \quad (4.7)$$

where $\text{Var}[y(t)] = \text{Var}[s_r(t)] + \text{Var}[s_h(t)] + \text{Var}[n(t)]$. $\rho_r(\tau)$, $\rho_h(\tau)$ and $\rho_n(\tau)$ denote the ACF of respiration, heartbeat and noise respectively. Since we have $\text{Var}[s_r(t)] \gg \text{Var}[s_h(t)]$ and $\text{Var}[s_r(t)] \gg \text{Var}[n(t)]$, we have the approximation that $\rho(\tau) \approx \rho_r(\tau)$. The ACF will have a definite peak at a certain delay which corresponds to the breathing cycle as shown in Fig. 4.4 (e).

4.3.2.3 Motion Detection

Random Body Motion (RBM) has been one of the most difficult technical challenges to wireless vital sign monitoring. Compared with the millimeter-scale chest movement caused by heartbeats, the scale of RBM can be tens of centimeters. The time-variant part of CIR phase measurement with RBM can be modelled as

$$y(t) = s_m(t) + s_r(t) + s_h(t) + n(t), \quad (4.8)$$

where $s_m(t)$ is the distance change caused by motion. Fig. 4.4 (c) shows an example of the phase measurement with motion. Note that when the scale of RBM is much larger than the respiration signal, the variation $\text{Var}[s_m(t)] \gg \text{Var}[s_r(t)] \gg$

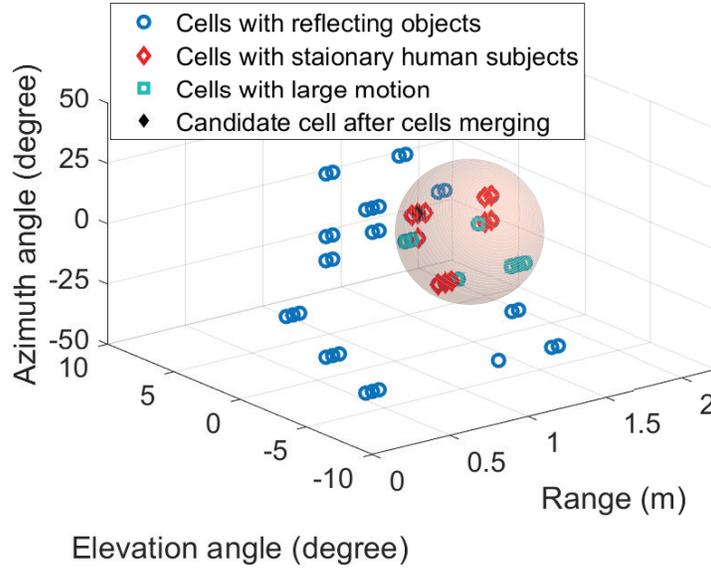


Figure 4.5: Example of cell merging.

$\text{Var}[s_h(t)]$, and thus $V_t(l, \theta, \phi) > \eta_{\max}$, where η_{\max} is a predefined threshold. When the subjects have moderate RBM, the variance of phase may be within the threshold, however, since RBM lacks periodicity in most case, we **cannot** observe a peak in $\rho(\tau)$ as the stationary case as shown in Fig 4.4 (f). Therefore, we have the motion indicator $\mathbf{1}_M(\cdot)$ defined as

$$\mathbf{1}_M(\theta, \phi, l) = \mathbb{1}(V_t(\theta, \phi, l) > \eta_{\max} \cup \rho(\tau_b) < \Gamma), \quad (4.9)$$

where $\mathbb{1}(\cdot)$ is the indicator function, τ_b is the first peak of ACF $\rho(\tau)$, and Γ is a predefined threshold.

4.3.3 Cell Merging/Clustering

Due to the fact that more than one cells have the RF signals reflected off a single human subject, a target clustering method should be employed before

determining the target number and vital sign monitoring. Considering the size of the human body, we can merge them into a cluster if the spacial distance between these cells is within the threshold d_{\min} . In ViMo system, we set the d_{\min} as 0.8m considering the typical body volume of human subjects. The cluster center of stationary cells is the cell with the largest ACF peak, corresponding to the cell with the human chest. The center of the RBM cells for each cluster is the cell with the largest $V_t(\theta, \phi, l)$, corresponding to the cell with the largest body motion. Note that even for a stationary person, he/she can have body motion from the body part away from the chest. So when the distance between stationary cells and RBM cells is smaller than the threshold d_{\min} , then these cells belong to the same person, and the center of the cluster should be the representative of stationary cells. The number of people is estimated by the number of clusters, where the location of each person is the center of its corresponding cluster.

Fig. 4.5 shows an example for cells merging, where the ground truth is that a human subject sits at 1m away from the device in a typical indoor office. The reflecting objects detected by the CFAR detector is shown in blue circles. The motion detector further differentiates cells with stationary subjects and RBM, shown as red diamonds and green squares respectively. The representative of the target is shown in solid black diamond.

4.4 Heart Rate Estimation

In ViMo, we enable the heart rate estimation module once a stationary subject has been detected. Since we check periodicity using ACF to determine whether the cell contains a stationary respiration signal as shown in Section 4.3.2.2, we can easily determine the breathing cycle by finding the peak location τ_r of $\rho(\tau)$, and the breathing rate should be

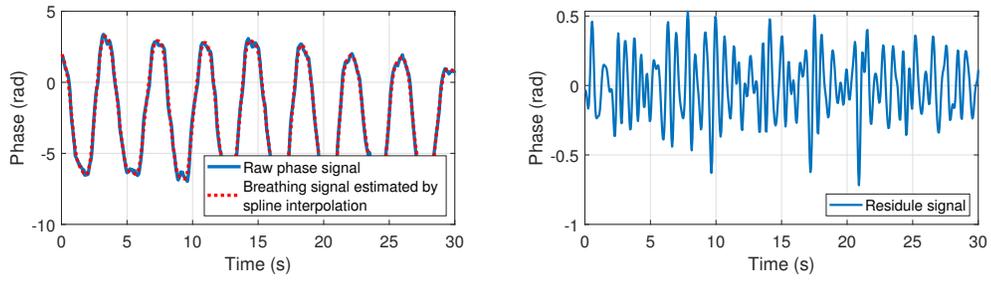
$$f_r = \frac{60}{\tau_r} \quad (4.10)$$

Respiration Per Minute (RPM).

Note that heartbeats introduce minute movements of the chest [57], which can be detected as small peaks in the unwrapped phase as shown in Fig. 4.6 (a). Past works [5, 8, 85] try to directly utilize frequency analysis and Band-Pass Filter (BPF) to estimate the heart rate. However, due to the harmonics introduced by respiration, it is easy to pick up the wrong peak for estimation as shown in the blue line in Fig. 4.6 (c). Thus, in order to get a higher estimation accuracy, we first eliminate breathing signal before heart rate estimation.

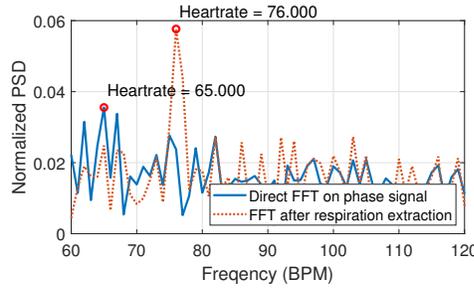
4.4.1 Breathing Interference Elimination

Eliminating the breathing signal can improve the Signal-to-Interference-plus-Noise Ratio (SINR) of the heartbeat signal, and thus improve the estimation accuracy. The polynomial fitting has been used [44, 50, 77] to remove respiration motion. However, one of the main drawbacks of the polynomial fitting is



(a) Breathing signal estimation.

(b) Residual after breathing elimination.



(c) PSD of original phase and its residual.

Figure 4.6: Example of breathing interference elimination.

the order selection. In previous works, the order is carefully selected by empirical experience, but under-fitting or over-fitting can be easily triggered when the experimental setting is changed (e.g., change of sampling rate or window length). Besides, the elimination effect is also related to the breathing rate. In other words, in order to achieve a similar elimination effect, the polynomial order should adapt to the user's breathing rate, which is not practical for robust daily deployment. To avoid this effect, ViMo adopts smoothing spline to estimate the breathing signal.

Let $\{t_i, y(t_i) : i = t_0, t_0 + T_s, \dots, t_0 + W\}$ to be a set of observations in the current window, where $T_s = \frac{1}{F_s}$ is the time interval between two adjacent samples, t_0 is the initial time of the observation window, and W is the window length. Compared

to the heartbeat signal, the respiration signals have larger distance change and lower frequency, thus, the estimate of the breathing signal $s_r(t)$ should be the solution of

$$\min_{\hat{f}} \sum_{i=t_0}^{t_0+W} \{y(t_i) - \hat{f}(t_i)\}^2 + \lambda \int \hat{f}''(t)^2 dt, \quad (4.11)$$

where $\lambda \geq 0$ is a smoothing parameter. The second term evaluates the smoothness of a function. The smoothing parameter controls the trade-off between fidelity to the data and smoothness of the function estimate. \hat{f} is the estimate of $s_r(t)$, defined as

$$\hat{f}(t) = \sum_{i=t_0}^{t_0+W} \hat{f}(t_i) f_i(t), \quad (4.12)$$

where $f_i(t)$ are a set of spline basis function [1].

To get the optimum solution of (4.11), we first define the vector $\hat{\mathbf{m}} = [\hat{f}(t_0), \dots, \hat{f}(t_0 + W)]^\top$, and the roughness penalty has the form

$$\int \hat{f}''(t)^2 dt = \hat{\mathbf{m}}^\top \mathbf{A} \hat{\mathbf{m}}, \quad (4.13)$$

where the elements of \mathbf{A} are $\int f_i''(t) f_j''(t) dt$. Therefore, the penalized sum-of-squared fitting error can be written as

$$\min_{\hat{\mathbf{m}}} \{\mathbf{y} - \hat{\mathbf{m}}\}^\top \{\mathbf{y} - \hat{\mathbf{m}}\} + \lambda \hat{\mathbf{m}}^\top \mathbf{A} \hat{\mathbf{m}}, \quad (4.14)$$

where $\mathbf{y} = (y(t_0), \dots, y(t_0 + W))^\top$. The minimizer of problem (4.14) is thus

$$\hat{\mathbf{m}}^* = (\mathbf{I} + \lambda \mathbf{A})^{-1} \mathbf{y}. \quad (4.15)$$

The heartbeat after elimination of the breathing signal is thus

$$\hat{s}_h(t) = y(t) - \hat{\mathbf{m}}^{*\top} \mathbf{f}(t), \quad (4.16)$$

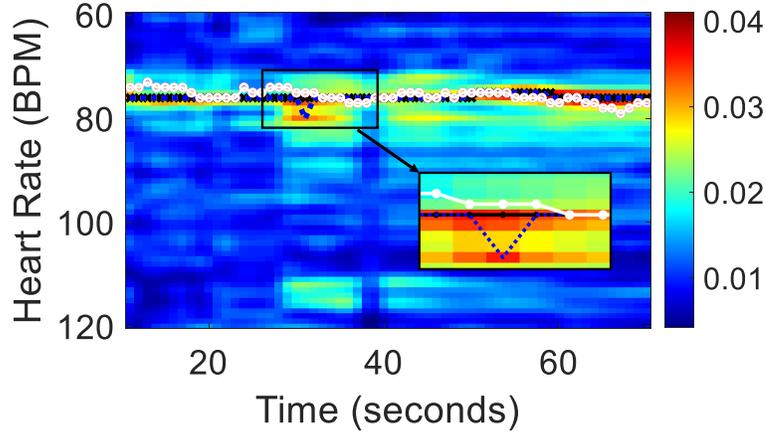


Figure 4.7: Spectrogram of residual signal.

where $\mathbf{f}(\mathbf{t})$ is the vector form of spline basis functions.

The dashed line in Fig. 4.6 (a) shows the estimation of the breathing signal. After breathing signal elimination, the spectrum of the residual signal after applying a BPF with passing band $[h_{\min} h_{\max}]$ is shown in the orange dashed line in Fig. 4.6 (c). The spectrum of the phase without eliminating respiration signal using the same BPF is shown in the blue solid line. It is obvious that the SINR of the heartbeat signal after breathing elimination is boosted. Specifically, the SINR is boosted from 1.65 dB to 5.65 dB by eliminating the respiration signal.

4.4.2 Heart Rate Estimation using Spectrogram

Breathing signal elimination can enhance the SINR of the heartbeat signal, and thus, increase the accuracy of heart rate estimation. However, the random measurement noises can still corrupt the estimation at some time instances. To further increase the estimation accuracy, in ViMo, we leverage the station-

ary property of heart rate and utilize the diversity in both frequency and time domains for reliable estimation.

Note that the heart rate can smoothly change over time, we model the heart rate as a Markov process, where the variation of heart rate between two adjacent time bins follows a normal distribution $\mathcal{N}(0, \sigma^2)$, and the Probability Density Function (PDF) is denoted as $p(f)$. After breathing signal elimination, we perform *Fast Fourier Transform* (FFT) on the residual and concatenate the PSD of each window to get a spectrogram as shown in Fig. 4.7.

Since the operation of FFT automatically discretizes the continuous frequency in the range of $[h_{\min}, h_{\max}]$ ² into $|\mathcal{Q}|$ frequency components, where $|\mathcal{Q}|$ means the cardinality of set \mathcal{Q} , the heart rate can be modelled as a Markov chain, and the transition probability matrix is denoted as $\mathbf{P} \in \mathbb{R}^{|\mathcal{Q}|} \times \mathbb{R}^{|\mathcal{Q}|}$, which is defined as

$$\begin{aligned} \mathbf{P}(q, q') &= \mathbf{P}(g(n) = q' | g(n-1) = q) \\ &= \int_{(q' - q - \frac{1}{2}) * \Delta f}^{(q' - q + \frac{1}{2}) * \Delta f} p(f) \mathbf{d}f, \end{aligned} \tag{4.17}$$

where $\forall q, q' \in \mathcal{Q}$. Here, $g : [1, N] \rightarrow \mathcal{Q}$ is a mapping indicating the frequency component at the given time, and N is the total time instances of a given spectrogram.

In principle, the heartbeat signal is more periodic than noise and other motion interference. Thus, it is more likely to be observed as peaks in most of the time. Moreover, considering that one's heart rate will not fluctuate a lot within a short period, estimations of heart rates should form a trace that achieves a good balance between frequency power and temporal smoothness.

²In ViMo, we set h_{\min} and h_{\max} as 60 Beat Per Minute (BPM) and 120BPM respectively.

The most probable heart rate trace can be found by solving

$$\mathbf{g}^* = \arg \max_{\mathbf{g}} E(\mathbf{g}) - \kappa C(\mathbf{g}), \quad (4.18)$$

where κ is a regularization factor. \mathbf{g} is denoted as a trace, where

$$\mathbf{g} = (n, g(n))_{n=1}^N. \quad (4.19)$$

$E(\mathbf{g})$ is the power of trace \mathbf{g} , defined as

$$E(\mathbf{g}) = \sum_{n=1}^N \mathcal{E}(n, g(n)), \quad (4.20)$$

where $\mathcal{E}(n, q)$ denotes the energy at time bin n and frequency component q . The smoothness of the trace can be evaluated by a cost function $C(\mathbf{g})$, defined as

$$C(\mathbf{g}) \triangleq -\log \mathbf{P}(g(1)) - \sum_{n=2}^N \log \mathbf{P}(g(n-1), g(n)), \quad (4.21)$$

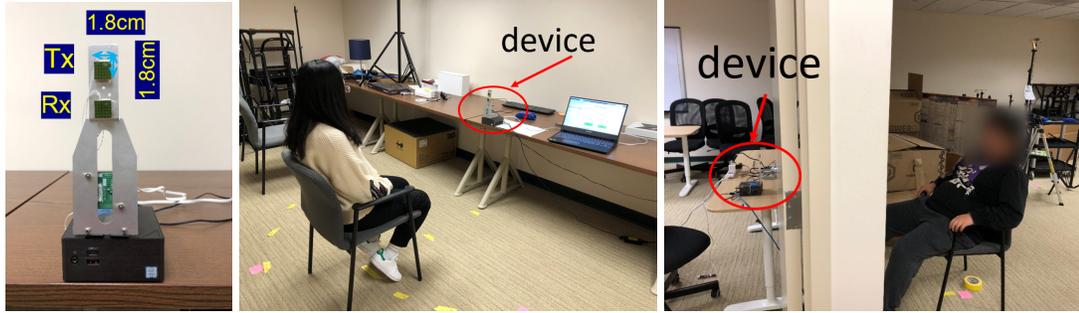
where the frequency transition probability $\mathbf{P}(g(n-1), g(n))$ can be calculated by (3.5). Without loss of generality, we assume a uniform prior distribution, i.e., $\mathbf{P}(g(1)) = \frac{1}{|\mathcal{Q}|}$.

This problem can be solved by dynamic programming [69]. The result of heart rate trace estimation is shown as the black line in Fig. 4.7, where the reference measured by a gold standard electrocardiogram (ECG) sensor [53] is marked as the white line. The estimation result of directly using the location of the highest peak of the spectrum as the heart rate estimation is shown as the blue dashed line. We can see that when the noise is too large (e.g., at time instance $t = 32$ s), the estimation without dynamic programming will locate at the wrong peak, resulting in a large estimation error. By using dynamic programming, the maximum estimation error decreases from 5 BPM to 3BPM, as shown in Fig. 4.7.

4.5 Experiment Evaluation

In this section, we evaluate ViMo in practical settings using a commodity 802.11ad chipset in a typical office of size $3.5 \text{ m} \times 3.2 \text{ m}$ as shown in Fig. 4.8. We embed ViMo in a Commodity Off-The-Shelf (COTS) 60GHz WiFi [55] as shown in Fig. 4.8 (a). Specifically, the chipset we used for ViMo has 32 antennas assembled in a 6×6 layout with a form factor of $1.8 \text{ cm} \times 1.8 \text{ cm}$ for both the transmitter (Tx) and the receiver (Rx). The chipset operates at 60GHz center frequency with 3.52GHz bandwidth, providing a range resolution of 4.26 cm. To extract CIR, the Tx transmits a known pulse composed of a complementary Golay sequence. A Golay correlator is implemented in the Rx hardware and the correlation result corresponds to the CIR. We enroll 8 participants (4 male and 4 female) aging from 22 to 35 for testing. The ground truth is provided by a commercial ECG sensor with a chest strap [53]. The sampling rate of ViMo is 20Hz.

To further evaluate our system, we compare it with the mmVital [85], which is the state-of-art wireless vital sign monitoring system using an impulse-based mmWave radio. mmVital leverages the RSS from a pair of horn antenna and finds the highest magnitude peak as well as its adjacent bins in the frequency domain to form a custom narrow BPF, and then counts peaks of the time-domain filtered signal to determine the breathing rates as well as heart rate. In order to make a fair comparison, same as ViMo, phases of CIRs from the detected cells are used as the input of mmVital algorithm, rather than the coarse information of RSS. To estimate both respiration and heart rate, the adaptive narrow BPF and



(a) Device

(b) LOS setting

(c) NLOS setting

Figure 4.8: Experiment setup.

IFFT are implemented as illustrated in mmVital [85]. The window length for both mmVital and ViMo are set to be 60 seconds, and systems give output every second. Note that mmVital [85] estimates vital signs according to the number of peaks in a time window (i.e., the estimation is an integer), the resolution of its breathing rate and heart rate estimation are 1 RPM and 1 BPM respectively.

4.5.1 Overall performance

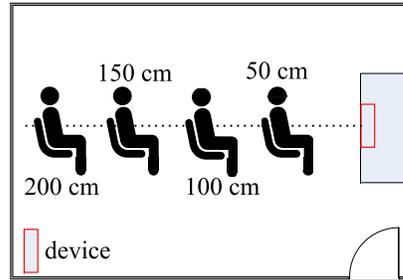
We first report the overall performance of ViMo in the measurement of Respiration Rate (RR) and Heart Rate (HR). The accuracy is calculated over 17 experimental runs of 3 minutes for all 8 participants. During the experiment, participants wear casual clothes, such as sweaters and shirts, sit in front of the device and breathe normally. Various factors including user and device placement diversity (e.g., distance, orientation, incident angle and blockage), motion interference and multiple users' position are considered. The detection rate of the system is 97.86% and the overall median error of RR and HR evaluated by

ViMo is 0.19 RPM and 0.92 BPM respectively. mmVital achieves similar performance w.r.t. RR, but its median error of HR is 1.6 BPM, 73.91% worse than ViMo. We further compare the performance of ViMo with Vital-Radio [5], which is the representative multi-user vital sign monitoring system build upon FMCW radar. Vital-Radio achieves 98.5% median accuracy of HR estimation (0.95 BPM) when the person is 1 m away from the device, whereas, the median error of HR estimation for ViMo is 0.58 BPM. Note that although the SNR of our device is lower than FMCW radar, ViMo still achieves better performance. The advantage of our system benefits from the breathing signal elimination module and Dynamic Programming (DP) algorithm when estimating the HR, which can increase the SINR of the heartbeat signal.

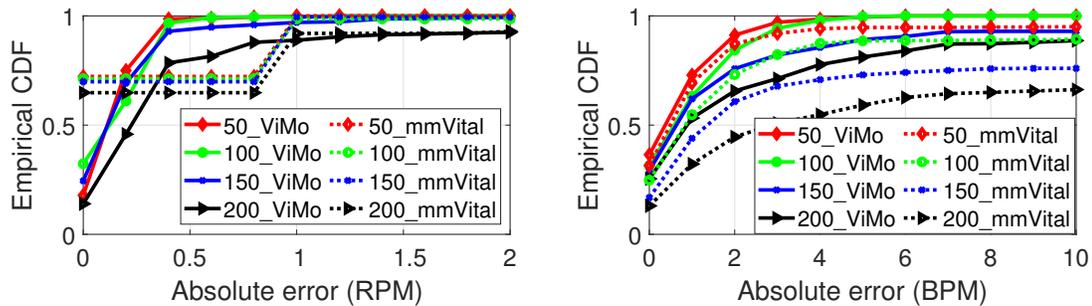
Moreover, experimental results show that ViMo can effectively detect the stationary/non-stationary state of human subjects, and can make accurate estimates of both RR and HR when slight user motion incurs (e.g., shaking head). Comparing with mmVital, which does not take users' motion into consideration, ViMo makes an important improvement towards practical deployment. The details will be discussed in the following sections.

4.5.2 Impact of Distance

In this section, we investigate the effect of the distance between the device and human subject on the estimation accuracy. Participants sit at different distances facing the device as shown in Fig. 4.9 (a). The empirical Cumulative



(a) Experiment setup



(b) CDF of absolute RR estimation error (c) CDF of absolute HR estimation error

Figure 4.9: Experiment setup and result for the impact of distance.

Distribution Function (CDF) of the absolute error of RR and HR estimation are shown in Fig. 4.9 (b) and Fig. 4.9 (c) respectively, where the performance of ViMo and mmVital are shown in solid lines and dash lines respectively. To account for the mis-detection, we set the estimation to be 0 BPM when the target is missed.

As expected, the performance degrades with distance due to the SNR degradation. The median error for RR of ViMo is within 0.15 RPM when the distance is within 1.5 m and it increases to 0.22 RPM when the distance increases to 2 m. For HR estimation, the median error of ViMo increases from 0.42 BPM to 0.9 BPM when the distance increases from 0.5 m to 2 m. Furthermore, we can see that the degradation of RR estimation is less than the HR estimation due to the

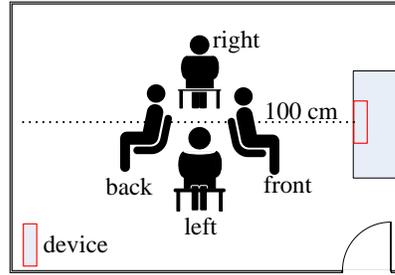
higher SNR of the breathing signal.

The CDF of RR estimation using mmVital algorithm is stepwise since the resolution of both ground truth and estimation is 1 BPM. It is obvious that both algorithms achieve similar performance as for RR estimation, but ViMo achieves a higher resolution. Moreover, for HR estimation, ViMo outperforms mmVital for all the 4 settings, and the performance gap becomes larger with the increment of distance. The main reason is that the breathing signal elimination helps to improve the SINR of the heartbeat signal as discussed in Section 4.4.1. Besides, DP algorithm in ViMo also leverages the time diversity besides the frequency diversity to make estimations, which can further alleviate the impact of the measurement noises.

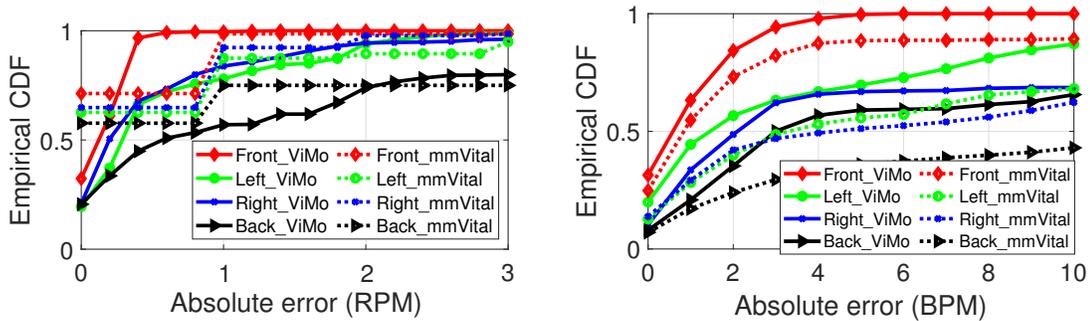
4.5.3 Impact of Orientation

In this study, we investigate the impact of human orientation on estimation accuracy. The orientation corresponds to the closest part of the user w.r.t. the device as shown in Fig. 4.10 (a). The distance from the user to the device is set to be 1 m. Fig. 4.10 (b) and Fig. 4.10 (c) show the estimation performance of RR and HR respectively.

It is shown that the “front” setting achieves the best performance, whereas, the “back” setting has the worst performance, for both RR and HR estimation. This result is due to the distinct displacement of reflecting part caused by respiration in different orientations. Since smaller displacement means lower SNR of



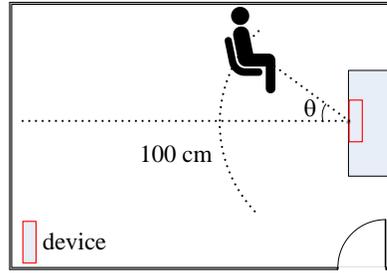
(a) Experiment setup



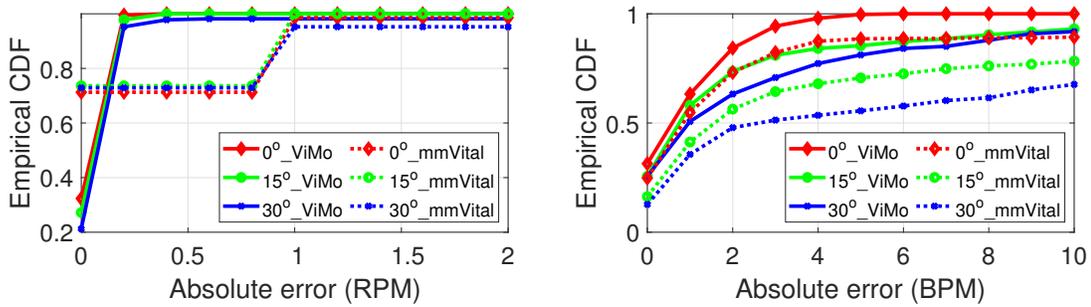
(b) CDF of absolute RR estimation error (c) CDF of absolute HR estimation error

Figure 4.10: Experiment setup and result for the impact of orientation.

breathing signal, when the displacement is too small, mis-detection occurs. The detection rate when the subject sits facing the device is 100%, and it degrades to 99.06% and 99.37% when the left and right side of the chest facing the device. The detection rate drops to 83.83% when human subjects sit back to the device. It is worth noting that even similar detection rates are achieved when participants sitting at the left and right orientation, the HR estimation performance is distinct, where the “left” setting outperforms the “right” setting. This is due to the physiological structure of human beings, where the vibration caused by the heartbeat is larger on the left side of the chest. Similarly, ViMo has equivalent performance in terms of RR estimation compared with mmVital, however, it has



(a) Experiment setup



(b) CDF of absolute RR estimation error (c) CDF of absolute HR estimation error

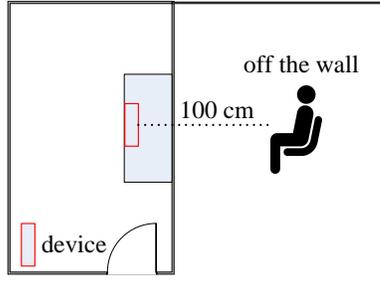
Figure 4.11: Experiment setup and result for the impact of incident angle.

much better performance of HR estimation for all the 4 settings, as shown in Fig. 4.10 (c).

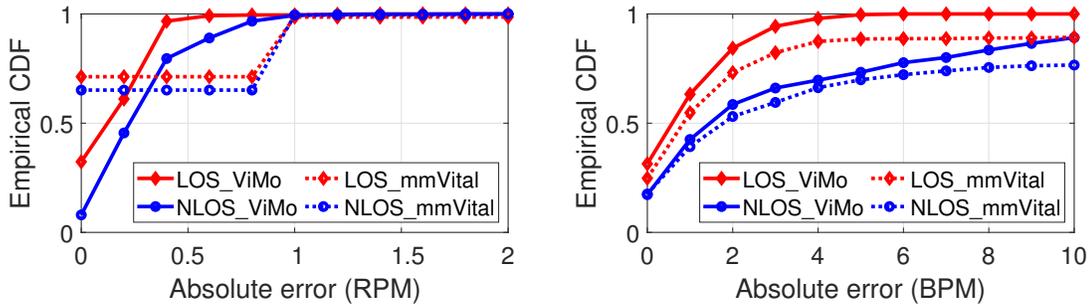
4.5.4 Impact of Incident Angle

In this part, we investigate the impact of the incident angle on the estimation performance, where human subjects are asked to sit at angles $[0^\circ, 15^\circ, 30^\circ]$, and the distance between human and device is 1 m, as shown in Fig. 4.11 (a). The CDF of the absolute estimation error of RR and HR with different incident angles are shown in Fig. 4.11 (b) and Fig. 4.11 (c) respectively.

We can see that for both RR and HR estimation, the accuracy decreases with



(a) Experiment setup



(b) CDF of absolute RR estimation error (c) CDF of absolute HR estimation error

Figure 4.12: Experiment setup and result for the impact of blockage.

the increment of the incident angle. The reason is that the reflection loss depends on the incident angle, and increment in incident angle increases the reflection loss, rendering lower SNR of the reflected signal. However, since the SNR of the breathing signal is much higher than the heartbeat signal, the performance degradation of RR estimation is not as severe as HR estimation. Furthermore, we can see that the performance of ViMo is much better compared with mmVital in terms of HR estimation, especially in the case of a large incident angle.



Figure 4.13: Experiment setup with different blockage materials.

4.5.5 NLOS Case

The RR and HR estimation accuracy are evaluated for the through-the-wall case, and the experiment setup is shown in Fig. 4.12 (a). Participants are asked to sit on the other side of a drywall, and the distance between the device and the human subject is 1 m. The median error of RR estimation increases from 0.15 RPM to 0.25 RPM due to the penetration loss, and the median error of HR estimation increases from 0.6 BPM to 1.4 BPM, as shown in Fig. 4.12 (b) and Fig. 4.12 (c) respectively.

In order to further investigate the influence of blocking material (corresponding to different penetration loss), we conduct a set of experiments, where different commonly used materials are used to block the LOS path, as shown in Fig. 4.13. Since the penetration loss is distinct for different material, the performance drop is different. The Mean Absolute Error (MAE) of RR and HR estimation is shown in Table. 4.1.

Besides, note that clothes can be considered as a cotton layer between the

Blocking material	None (LOS)	Glass	Cotton pad	Wood panel	White board	Drywall
Mean RR error (RPM)	0.14	0.23	0.24	0.26	0.28	0.29
Mean HR error (BPM)	1.29	2.66	3.45	4.82	4.85	5.95

Table 4.1: Performance with different blockage materials.

Clothes Type	T-shirt	Sweater	Winter jacket
Median absolute error of RR (RPM)	0.125	0.142	0.164
Median absolute error of HR (BPM)	0.5	0.58	1.13

Table 4.2: Impact of thickness of clothes.

human subject and the device, and the penetration loss is related to the thickness of clothes. To investigate the influence of the thickness of clothes, we conduct a similar experiment as that in Fig. 4.13, where participants are asked to wear T-shirts, sweaters and winter jackets during the experiment. The median absolute error of RR and HR are shown in Table. 4.2. We can see that the estimation accuracy decreases with the increment of the thickness of clothes. However, for the general indoor wearing (e.g., T-shirt and sweater), the degradation can be negligible.

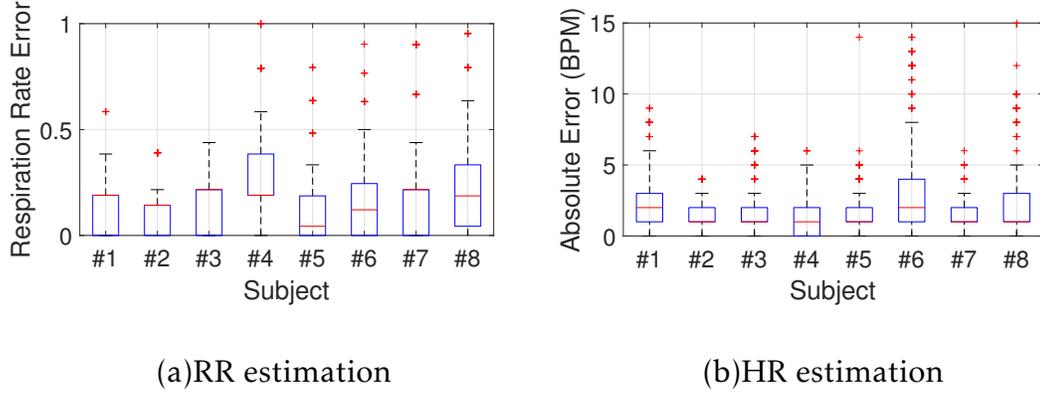


Figure 4.14: Impact of user heterogeneity on estimation accuracy.

4.5.6 Impact of User Heterogeneity

In this part, we investigate the impact of the user heterogeneity on the performance. The data of all the settings above are used to get the performance of each user. The difference in error distribution can be caused by various factors, such as reflection loss and heartbeat strength, etc. Fig. 4.14 (a) gives the RR estimation performance, with the maximum median error within 0.2 RPM for all participants. Fig. 4.14 (b) shows the error distribution of absolute HR error of all 8 subjects, where all of them have a median error within 2 BPM.

4.5.7 Impact of Body Movement

We evaluate the performance of ViMo when users have different motion states. Participants are asked to shake head (1 ~ 3 cm) and move the body (4 ~ 5 cm) every 20 seconds. The distance from the device to the user is 1 m with incident angle 0° . We also evaluate the performance when users answer phone with headset (continuously talking). The MAE of RR and HR are shown in Table. 4.3.

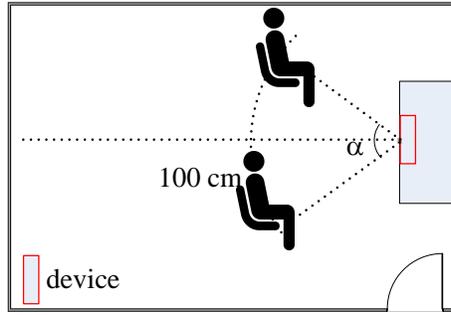
Body state	Stationary	Shaking head (left-right)	Moving body (left-right)	Speaking
Mean RR error (RPM)	0.14	0.28	0.51	1.22
Mean HR error (BPM)	1.29	4.16	3.06	6.31

Table 4.3: Performance for different motion states.

All the cases achieve more than 99.7% detection rate, where for the case of moving body, in 27% of the duration we detect large body motion, and thus the vital signs estimation module will not be triggered. As for the time that body motion is within the detection threshold (a.k.a, stationary period), the vital sign estimation module is triggered, and the mean HR error is 3.06 BPM for the case of moving body (the relative error is 4%). However, for the case when people are answering the phone, the chest will involve RBM caused by speaking more frequently, resulting in the worst performance for all the test cases.

4.5.8 Multi-user Case

In this part, we first study the impact of the angular separation between users, where two users sit at a distance of 1 m away from the device with different separation angles as shown in Fig. 4.15. We define the Detection Index (DI) of a separation angle as the ratio between the number of samples when the number of detected targets matches the ground truth and the total number of samples. We



(a) Illustrate



(b) Experiment setup

Figure 4.15: Experiment setup for the impact of separation angle.

also define the False-alarm Index (FI) of a separation angle as the ratio between the number of samples when the number of detected targets is larger than the ground truth and the total number of samples. Table. 4.4 shows the median error of RR and HR estimation for both users.

Compared to the single-user scenario, the performance degrades at small separation angles (i.e., 30°), but the performance is similar to the single-user scenario if the separation angles are large enough (i.e., larger than 45°). This is because when the distance of two targets is small enough, the distance of the candidate cells with each user can be smaller than the predefined threshold d_{\min} . Thus, the two clusters will be merged together and there will be only one representative cell left, resulting in a mis-detection. Besides, the cells with high SNR signals of one user can be merged with the other user's, therefore, the SNR of the representative cell for vital signs estimation can drop, resulting in degradation of the performance.

To further evaluate ViMo's accuracy for multi-user vital sign estimation,

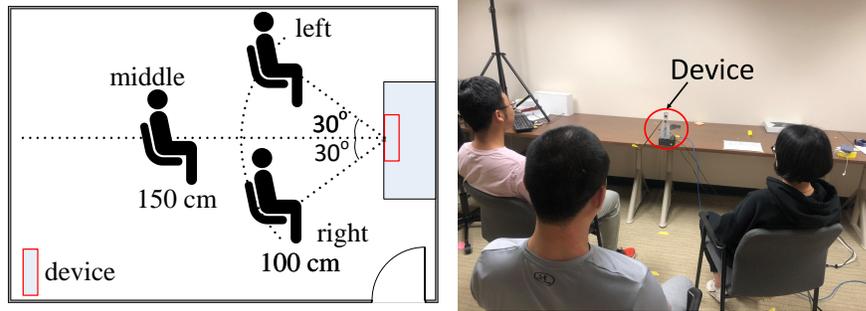
Separation angle	DI	FI	Med. error of breathing	Med. error of heart rate
30°	0.84	0	(1.14;0.15)	(2;1)
45°	0.98	0	(0.22;0.14)	(1;1)
60°	1	0	(0.21;0.14)	(1;1)
75°	1	0	(0.21;0.12)	(1;1)

Table 4.4: Performance for different separation angles.

we perform controlled experiments, where we ask 3 users to sit in parallel as shown in Fig. 4.16. ViMo detects the location of each target and simultaneously estimate their vital signs. When mis-detection happens, we define the relative error as 1 as before. Fig. 4.17 shows the mean relative accuracy of RR and HR as well as the detection rate at each location. We can see that for all the 3 locations, ViMo achieves the mean accuracy of both RR and HR over 92.8%. As for the detection rate, since the separations between the middle location and the other two locations are not large enough, and the middle location is more distant, the detection rate drops at the middle location. However, the overall detection rate over time is still above 92.7% during the testing.

4.6 Summary

In this chapter, we present ViMo, a multi-person **Vital sign Monitoring** system using a single commercial 802.11ad device. We devise a multi-object stationary/non-stationary detection algorithm to locate and count human tar-



(a) Illustrate

(b) Experiment setup

Figure 4.16: Multi-user experiment setup.

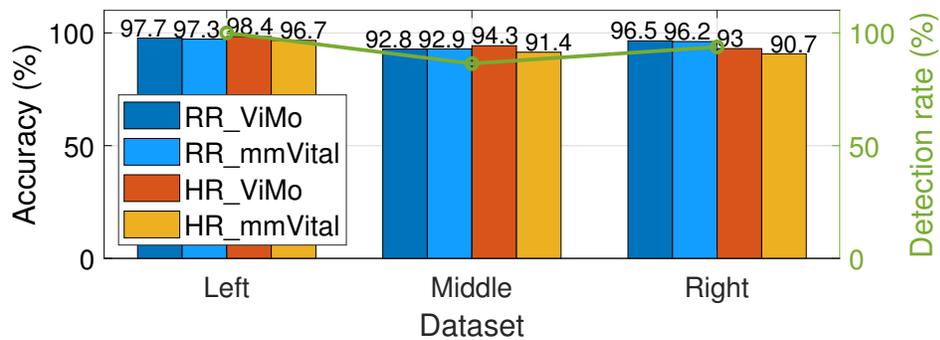


Figure 4.17: Multi-user accuracy and detection performance.

gets without any prior calibration. In addition to the instantaneous estimating breathing rates using ACF with high accuracy, we further design a robust heart rate estimator, which eliminates the interference of the breathing signal and then estimates the heart rate leveraging both the time and frequency diversity. We evaluate the performance of ViMo by various settings, including NLOS and motion artifacts, the most challenging scenarios for wireless vital signs monitoring. Experiment results show that ViMo can accurately monitor vital signs, with a median error of 0.19 RPM and 0.92 BPM, respectively, for RR and HR estimation.

Chapter 5: mmHRV: Contactless Heart Rate Variability Monitoring

5.1 Introduction

Heart Rate Variability (HRV), defined as the variation of the periods between consecutive heartbeats, i.e., Inter-Beat Intervals (IBI), is an important indicator of the overall health status of an individual [4]. Analysis of the HRV has been proved to be a powerful tool to assess cardiac health and evaluate the state of the Autonomic Nervous System (ANS) [64]. High-accuracy HRV monitoring is required in numerous applications such as early diagnose of cardiovascular disease, stress evaluation, emotions recognition and anxiety treatment, etc. [10, 17, 23, 68, 93].

Traditional measurements of the HRV are obtained by continuously measuring the IBIs using the electrocardiogram (ECG) or photoplethysmogram (PPG) sensors, both of which are dedicated medical devices and have to be physically contacted with the human skin. However, using ECG or PPG is uncomfortable for users and sometimes may cause skin allergies. Therefore, how to monitor the HRV in a non-contact way has become an important topic for both academia and

industry.

While many existing works have validated the feasibility of vital sign monitoring using RF signal, most of them can only estimate the Respiration Rate (RR) [5, 11, 91] and the Heart Rate (HR) [6, 45, 48, 75, 85], from which one cannot obtain the HRV without the precise timing of each heartbeat. As a result, accurate RF-based HRV monitoring needs to be further investigated.

In this chapter, we present mmHRV, the first multi-person HRV estimation system using Commodity Off-The-Shelf (COTS) millimeter-Wave (mmWave) radio. The rest of the paper is organized as follows. The system overview is presented in Section 5.2, followed by the target detection in Section 5.3 and the heartbeat extraction and HRV estimation in Section 5.4. The performance is evaluated in Section 5.5. We summarize the chapter in Section 5.6.

5.2 System Overview

mmHRV is a wireless system that can accurately detect the heartbeat signal of human subjects and estimate their HRV by purely using the RF signals reflected off the users' bodies. The processing flow of mmHRV is shown in Fig. 5.1. First, a Frequency-Modulated Continuous Wave (FMCW) radar transmits the RF signal and captures the reflections of human subjects and static objects. In order to detect human subjects at different locations, the *Bartlett beamformer* is applied to get the channel information at different azimuth-range bins. Then, we devise a *target detector* that adaptively estimates the noise level at various

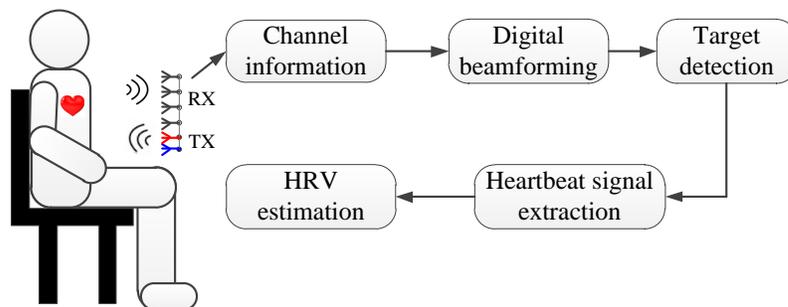


Figure 5.1: Processing flow of mmHRV.

distances and azimuth angles and thus detects the presence of reflecting objects. The variance of phase is further utilized to distinguish human subjects and static objects. To identify the number of target and their locations, a non-parametric clustering algorithm is employed.

To extract the heartbeat signal from the phase information that is modulated by both respiration and heartbeat, we devise a *heartbeat signal extractor*, which can decompose the phase signal into several narrow-band signals concurrently and give an estimate of heartbeat wave. HRV can be further analyzed based on the Inter-Beat Intervals (IBIs) derived from the estimated heartbeat signals.

5.3 Target Detection

For practical application, target detection needs to be performed before vital sign detection, which has been omitted in many works. The target detection is hard to achieve, especially in the indoor scenario, where there are various objects (e.g., wall, desk, metal objects, etc.) with strong reflections of EM waves.

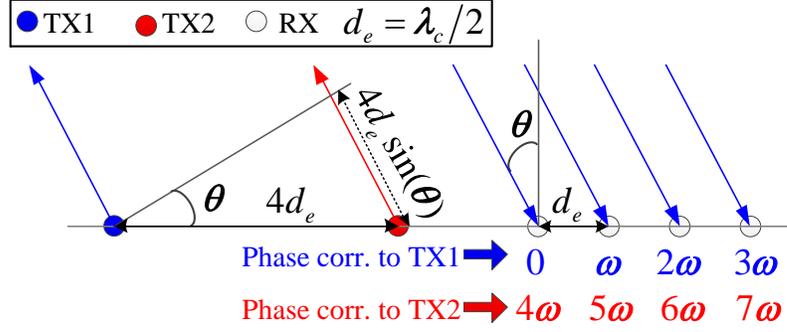


Figure 5.2: Antenna Deployment.

5.3.1 Range-FFT and Digital Beamforming

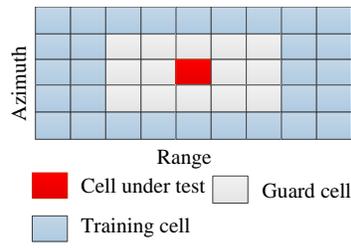
In our system, we take advantage of the multiple antennas of the chipset, and use 2 Tx antennas and 4 Rx antennas, as shown in Fig. 5.2. To increase the azimuth resolution, the chirps are transmitted in the Time-Division Multiplexing (TDM) mode [46] by transmitting sequentially through two Tx antennas. This is equivalent to the 8-element virtual array as shown in Fig. 5.2.

The channel information for the case when there is a static object is

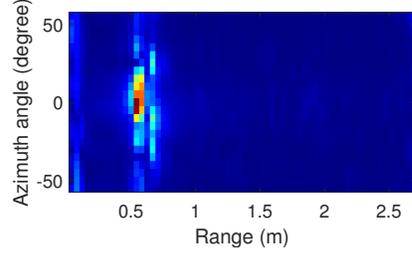
$$h(l, n, m) = A \exp \left\{ -j \left(2\pi f_b n T_f + \frac{4\pi d_0}{\lambda_c} + 2\pi \frac{d_l \sin \theta}{\lambda_c} \right) \right\}, \quad (5.1)$$

where d_l is the relative distance introduced by virtual antenna l . θ is the azimuth angle of the target as shown in Fig. 5.2. d_0 is the distance between the object and the device, which stays constant in slow time.

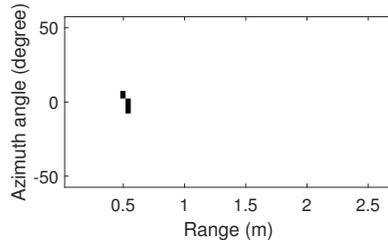
Note that the channel information corresponding to the reflecting object is a periodic signal in fast time, and the periodicity is related to the distance as shown in Equ. (2.15) and Equ. (5.1). To determine the range information of reflecting objects, the FFT is performed over the fast time for each chirp, i.e.,



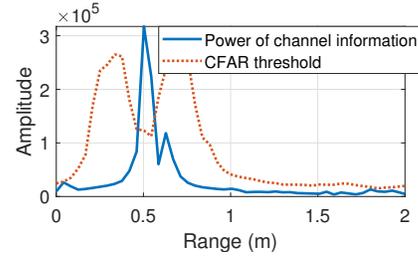
(a) 2D CFAR window.



(b) Amplitude information in the range-azimuth plane



(c) Result of the Reflecting object detector



(d) Example of the CFAR detector in range domain

Figure 5.3: Example of the reflecting object detector.

range-FFT, and the channel information can be written as $h_r(l, m)$, where r is the range tap index. The range taps corresponding to the reflecting objects would observe larger energy compared with that without reflecting objects. To further determine the azimuth angles of the reflecting objects, digital beamforming is performed over all antenna elements for each range tap, and the channel information corresponding to range r and azimuth angle θ can be expressed as

$$h_{r,\theta}(m) = \mathbf{s}^H(\theta)\mathbf{h}_{r,1}(m) + \epsilon(m), \quad (5.2)$$

where $\mathbf{s}^H(\theta)$ is the steering vector towards angle θ . In mmHRV, Bartlett beam-

former [40] is adopted, where the coefficient of the l -th antenna is

$$s_l(\theta) = \exp(-j2\pi \frac{d_l \sin \theta}{\lambda_c}). \quad (5.3)$$

$\epsilon(m)$ is the additive white Gaussian noise assumed to be independent and identically distributed (I.I.D) for different range-azimuth bins. the channel information vector at range tap r overall all antenna elements is denoted as $\mathbf{h}_{r,1}(m) = [h_{r,1}(m), h_{r,2}(m), \dots, h_{r,L}(m)]$. Therefore, for each sample m in slow time, we will have a channel information matrix $h(r, \theta)$, which contains channel information at different location bins with range r and azimuth angle θ . Fig. 5.3 (b) shows the amplitude of the channel information at the range-azimuth plane.

5.3.2 Reflecting Object Detector

To locate human subjects, we first need to identify the range-angle bins with reflecting objects. Note that the channel information for the bins without any reflecting object only contains noise, and thus, the energy of channel information for the bins with reflecting objects is larger than those without any reflecting objects, as shown in Equ. (2.15) and Equ. (5.1) respectively. However, it is impossible to find a universal predefined threshold for target detection. According to the propagation laws of EM wave, for the same reflecting objects, a shorter distance corresponding to a larger reflecting energy. In mmHRV, we utilize the Constant False Alarm Rate (CFAR) [60] detector, which can estimate the noise level by convolving the CFAR window (shown in Fig. 5.3 (a)) with the channel information at the range-azimuth plane (shown in Fig. 5.3 (b)), and the location

bins with reflecting objects are those whose energy is above the noise level, as shown in Fig. 5.3 (c). Fig. 5.3 (d) shows the example of CFAR detection in the range domain, where the threshold is shown in the dashed line.

5.3.3 Human Subjects Detector

Although *Reflecting object detector* can filter out the empty taps, it cannot distinguish human subjects from static reflecting objects. Note that different from static objects, the distance between human subjects and the device will change over slow time due to motions (e.g., respiration and heartbeat), and thus result in a phase change as shown in Fig. 5.4 (a). Therefore, to further filter out the static reflecting objects, we leverage the phase information of the candidate bins selected by the *Reflecting object detector*.

In specific, when the EM wave is reflected by a human subject, the phase will change over slow time due to the modulation of human motions. Therefore, there is a large phase variance for the bins corresponding to human subjects. However, for bins corresponding to the static objects (e.g. desk, wall, etc.), the phase variance will be much smaller, as shown in Fig. 5.4 (a) and Fig. 5.4 (b). So in mmHRV, to filter out the static objects, we check the variance of the phase information over slow time, and the bins corresponding to a human subject are those whose phase variance above a certain threshold.

Note that there will be more than one bin corresponding to a human subject considering the volume of a human subject, as shown in Fig. 5.4 (c). To iden-

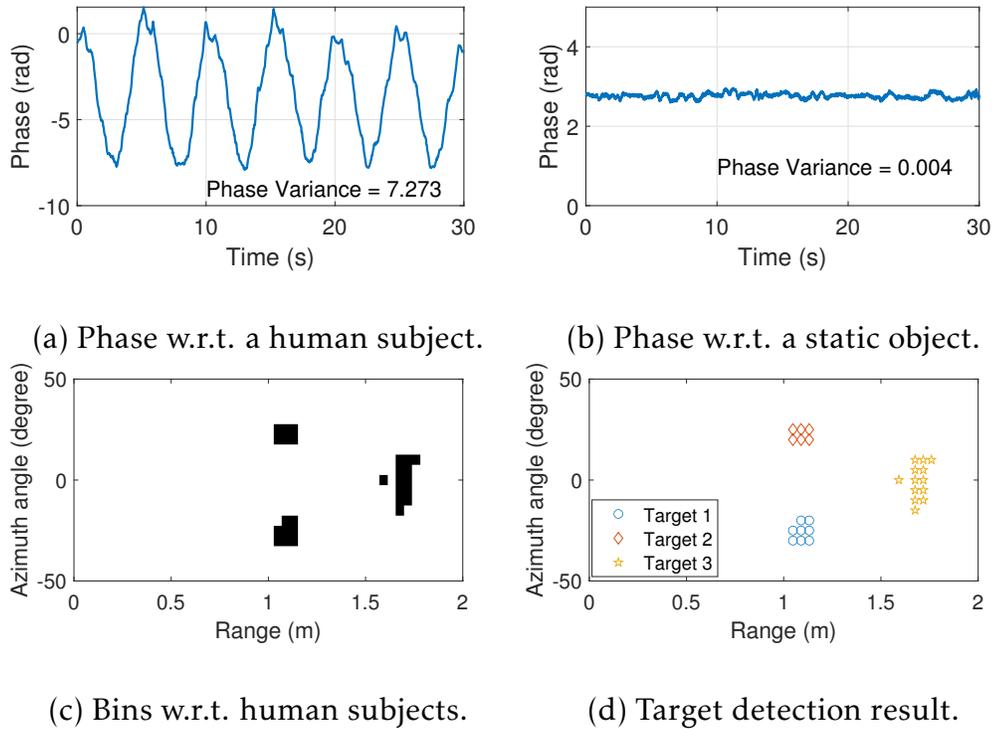


Figure 5.4: Example of the human subject detector.

tity the target number, mmHRV utilizes a non-parametric clustering method, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, to cluster the candidate bins without prior knowledge of cluster number. The clustering result is shown in Fig. 5.4 (d). The representative of each cluster is the bin with the best periodicity. In specific, the bin with the highest peak for the first peak of the auto-correlation is selected, which corresponds to the bin with the highest SNR of the vital signs [75].

5.4 Heartbeat Extraction and HRV Estimation

Estimating HRV requires accurate estimation of Inter-Beat Intervals (IBIs), therefore, we need to extract the displacement change caused by heartbeats (a.k.a.,

heartbeat wave) from the compound displacement change of chest wall and detect moments in which heartbeats occur.

5.4.1 Heartbeat Extraction Algorithm

5.4.1.1 Problem Formulation

Recall that the phase information reflects the distance change caused by vital signs. For simplicity, we directly use the analog form of signals, and the distance change of the human chest can be written as

$$y(t) = s_m(t) + s_r(t) + s_h(t) + n(t), \quad (5.4)$$

where $s_m(t)$ denotes the distance change caused by body motion. $s_r(t)$ and $s_h(t)$ denote the distance change caused by respiration and heartbeat, respectively. $n(t)$ is the random phase offset introduced by noise, which is independent with the phase change caused by vital signs.

Note that both $s_r(t)$ and $s_h(t)$ are quasi-periodic signals, where the period can slightly change over time. Besides, we assume the body motion introduces few oscillations, i.e., a base-band signal. Thus, the signals related with the human subject are sparse in the spectral domain and we can reconstruct these signals with a few band-limited signals. In specific, each component $u_k(t)$ is assumed to be compact around a center pulsation ω_k , which is to be determined along with the decomposition. Moreover, the decomposition should achieve the spectrum

sparsity and data fidelity at the same time, which is modeled as

$$\min_{u_k \in \mathcal{U}, \omega_k \in \Omega} \alpha \sum_{k=1}^K \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \exp(-j\omega_k t) \right\|_2^2 + \left\| y(t) - \sum_{k=1}^K u_k(t) \right\|_2^2, \quad (5.5)$$

where the first term evaluates the bandwidth of the analytic signal associated with each component, and the second term evaluates the data fidelity. K is the total number of decomposition components, where $\mathcal{U} = \{u_1(t), \dots, u_K(t)\}$ and $\Omega = \{\omega_1, \dots, \omega_K\}$ are the set for all components and their center frequencies, respectively. α is a parameter for balancing the bandwidth constraint and data fidelity.

Once the hyper-parameters are known, the optimization problem in (5.5) can be solved by alternatively updating $u_k(t)$ and ω_k until convergence [16]. To update u_k , the subproblem can be written as

$$u_k(t) = \arg \min_{u_k(t)} \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \exp(-j\omega_k t) \right\|_2^2 + \left\| y(t) - \sum_{i=1}^K u_i(t) \right\|_2^2. \quad (5.6)$$

By using the Parseval theorem, the problem can be rewritten as

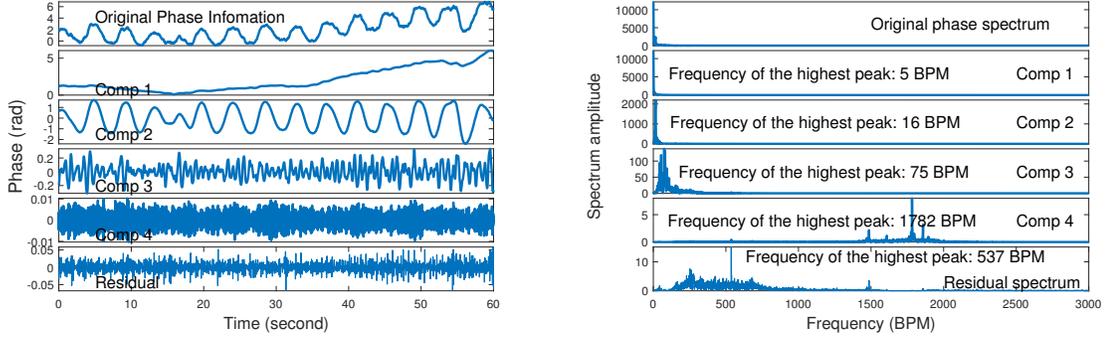
$$\square_k(\omega) = \arg \min_{\square_k(\omega)} \alpha \left\| j\omega \left[(1 + \text{sgn}(\omega + \omega_k)) \square_k(\omega) \right] \right\|_2^2 + \left\| \dagger(\omega) - \sum_{i=1}^K \square_i(\omega) \right\|_2^2, \quad (5.7)$$

where $\square_k(\omega)$ and $\dagger(\omega)$ are the Fourier transfer of $u_k(t)$ and $y(t)$ respectively. After taking integrals over frequency and performing a change of variable, we can get the updating formula, where

$$\square_k(\omega) = \frac{\dagger(\omega) - \sum_{i, i \neq k} \square_i(\omega)}{1 + 2\alpha(\omega - \omega_k)^2}. \quad (5.8)$$

Note that the center frequencies ω_k only appears in the bandwidth constraint and thus the subproblem can be written as

$$\omega_k = \arg \min_{\omega_k} \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \exp(-j\omega_k t) \right\|_2^2. \quad (5.9)$$



(a) Decomposition of a typical phase signal in time domain. (b) Corresponding spectrum of the decomposed components.

Figure 5.5: Example of heartbeat extractor.

As before, we find the optimum in Fourier domain, and we have

$$\omega_k = \arg \min_{\omega_k} \int_0^{\infty} (\omega - \omega_k)^2 |\Gamma_k(\omega)|^2 d\omega. \quad (5.10)$$

The minimizer of the above quadratic problem is

$$\omega_k = \frac{\int_0^{\infty} \omega |\Gamma_k(\omega)|^2 d\omega}{\int_0^{\infty} |\Gamma_k(\omega)|^2 d\omega}. \quad (5.11)$$

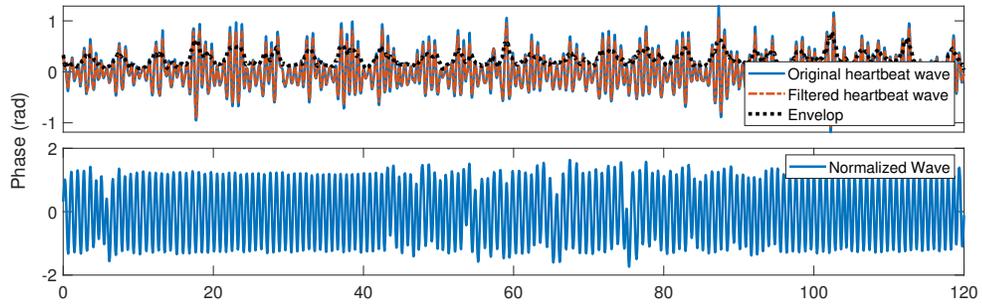
Fig. 5.5 illustrates the decomposition of a typical one-minute phase signal from the experiment, where the original phase information has been decomposed into 4 components. The first component reflects the body motion of the human subject, the second component is the respiration motion, and the third component is the heartbeat wave. Since the noise has different vibration characteristics as vital signals, it falls into a different mode as well as in the residual of the decomposition of the signal, as shown in Fig. 5.5.

5.4.1.2 Algorithm Design

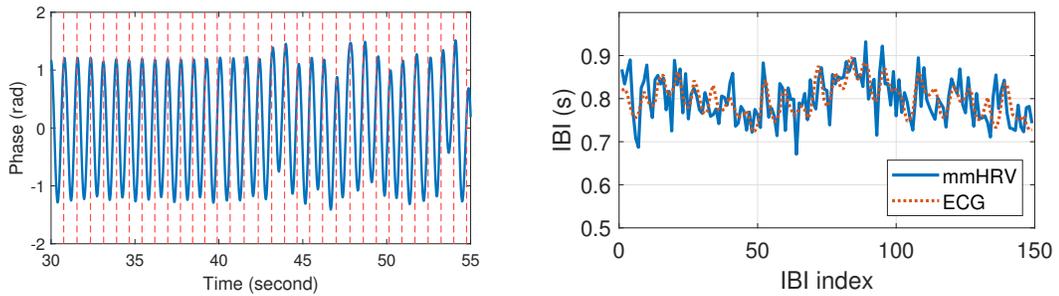
It has been proven that the decomposition problem can be solved once the hyper-parameters are properly defined. However, it is hard to predefine these hyper-parameters in real applications for heartbeat wave extraction. First, the human motion does not always exist and the human respiration sometimes will have a strong second harmonic component, making it even harder to determine the component number. Furthermore, the hyper-parameter α also influences the decomposition performance. Before discussing how to choose the hyper-parameter, we first discuss their influence on the decomposition result.

In specific, for the case that α is too small, i.e., the bandwidth constraint is too loose, when K is too small, the mixing problem will happen so that two signals may merge to a single decomposed component. However, when K is too large, some of the decomposed components may consist of noise. For the case that α is too large, i.e., the bandwidth constraint is too tight, when K is too small, some target signals may be discarded in noise. However, when K is too large, some important parts of the signal may be separated into two or more decomposed components.

In mmHRV, to accurately decompose the signal and get the component we are interested, i.e., the heartbeat wave, we are trying to adaptively change the component number K and α for different datasets. Here, we introduce a heuristic method to change K and α as the iteration proceeds to get proper decomposition result. Since the distance change caused by heartbeat is much smaller than the



(a) Heartbeat wave normalization.



(b) Normalized heartbeat wave v.s. ECG ground truth. (c) IBI estimation of mmHRV v.s. ECG ground truth.

Figure 5.6: Example of IBI estimation.

distance change caused by respiration and human motion, once the component corresponding to the heartbeat is decomposed, the component corresponding to respiration and motion should be decomposed as well, considering the data fidelity constraint in the objective function. Therefore, the algorithm will terminate once we get the component corresponding to the heartbeat. The details about the algorithm are shown in Algorithm 2.

Algorithm 2 Heartbeat wave extraction algorithm

```
1: Input  $y(t)$ 
2: Set  $\alpha \leftarrow \alpha_{\min}$ ,  $K \leftarrow K_{\min}$ 
3: repeat
4:   repeat
5:     Initialize  $\mathcal{U}$  and  $\Omega$ ,  $\text{flag} \leftarrow 0$ ,  $n \leftarrow 0$ 
6:     repeat
7:        $n \leftarrow n + 1$ 
8:       for  $k = 1 : K$  do
9:         update  $\Pi_k(\omega)$  using equ.(5.8); update  $\omega_k$  using equ.(5.11)
10:      end for
11:      until convergence:  $\sum_{k=1}^K \|\Pi_k^{n+1} - \Pi_k^n\|_2^2 / \|\Pi_k^n\| < \epsilon$  or  $n > n_{\max}$ 
12:      if exist  $\omega_k \in [h_{\min}, h_{\max}]$  and  $\text{Range}(\text{IFFT}(\Pi_k(\omega))) < r_{\max}$  then
13:         $\text{flag} \leftarrow 1$ , break;
14:      else
15:         $K \leftarrow K + 1$ 
16:      end if
17:      until  $K > K_{\max}$ 
18:      if  $\text{flag} == 1$  then
19:        break;
20:      else
21:         $\alpha = 2\alpha$ 
22:      end if
23: until  $\alpha > \alpha_{\max}$ 
```

5.4.2 HRV Estimation

Once the heartbeat wave is extracted, the exact time corresponding to each heartbeat can be identified by the peaks of the heartbeat wave. To further increase the accuracy, normalization is performed before peak extraction.

In specific, the envelope of the heartbeat wave is estimated by taking moving average to the absolute value of the heartbeat component, shown as the dashed line in Fig. 5.6 (a). We further perform a moving average filter to the original heartbeat wave to reduce the noise. The normalized wave is the ratio between the filtered heartbeat wave and the estimated envelope. IBIs can thus be derived by calculating the time duration between two adjacent heartbeats. Fig. 5.6 (b) shows a segment of heartbeat wave and its ECG ground-truth, where the dashed lines show the exact time of each heartbeat from a commercial ECG sensor [25]. The peaks of normalized heartbeat wave match with the ground-truth, and Fig. 5.6 (c) shows the estimated IBIs and the ECG ground-truth.

The HRV features can be further obtained from the IBI sequence. In mmHRV, we use the three most widely used metrics to evaluate the HRV [64]. One is the Root Mean Square of Successive Differences (RMSSD), which measures the successive IBI changes, and can be calculated by

$$\text{RMSSD} = \sqrt{\frac{1}{N_{\text{IBI}} - 1} \sum_{i=2}^{N_{\text{IBI}}} (\text{IBI}(i) - \text{IBI}(i - 1))^2}, \quad (5.12)$$

where N_{IBI} is the total number of IBIs of the measurement. The standard deviation of all the IBIs (SDRR) measures the variation of the IBIs, which can be

calculated as

$$\text{SDRR} = \sqrt{\frac{1}{N_{\text{IBI}}} \sum_{i=1}^{N_{\text{IBI}}} (\text{IBI}(i) - \overline{\text{IBI}})^2}, \quad (5.13)$$

where $\overline{\text{IBI}}$ is the empirical mean of the IBIs of each measurement. The metric pNN50 measures the percentage of successive IBI that differ by more than 50 milliseconds (ms), which can be calculated by

$$\text{pNN50} = \frac{\sum_{i=2}^{N_{\text{IBI}}} \mathbb{1}\{(\text{IBI}(i) - \text{IBI}(i-1)) > 50\text{ms}\}}{N_{\text{IBI}} - 1}, \quad (5.14)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

5.5 Experiment Evaluation

This section introduces the evaluation details of the proposed mmHRV, including practical system implementation, experiment setup, performance analysis and also the comparison with the state-of-art work.

5.5.1 Methodology

We prototype the mmHRV system by leveraging a commodity mmWave FMCW radar [26] in a typical office of size 3.5 m × 3.2 m as shown in Fig. 5.7. By configuring the 2 Tx antennas and 4 Rx antennas into TDM-MIMO mode as introduced in Section 5.3.1, the system can achieve a theoretical azimuth resolution of 15°. The Field of View (FoV) is 100° in the horizontal plane with a radius of about 4m [6], which is sufficient to cover typical rooms. To get the true heartbeat signal, an ECG sensor [25] (shown in Fig. 5.7 (a)) is used to collect the ground-truth simultaneously with the mmHRV during the experiment.

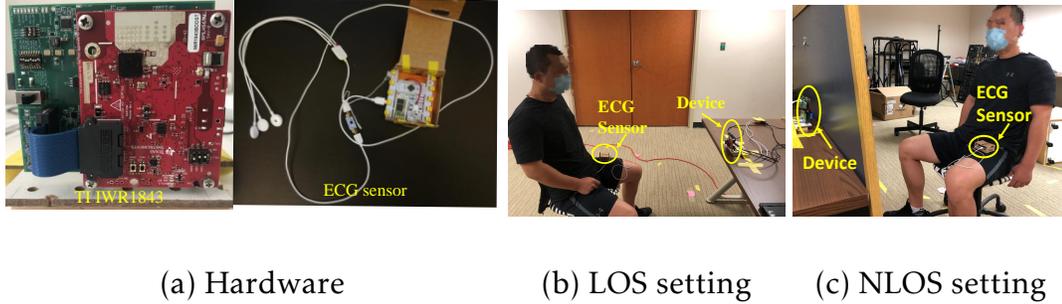


Figure 5.7: Experiment setup.

In total, 11 participants (6 males and 5 females) aging from 20 to 60 are invited to conduct experiments in both LOS and NLOS scenario as shown in Fig. 5.7 (b) and Fig. 5.7 (c). We conduct the experiments with a variety of settings including different distances, incidental angles, orientations and blockages between the human subject and the radar.

To further evaluate the performance of the proposed system, we compare mmHRV with the state-of-the-art HRV estimation technique using Band-Pass-Filter-Bank (BPFB) [52], where the BPF is used to eliminate respiration interference before heartbeat wave estimation.

The HR is then estimated and the heartbeat signal is estimated by using the narrow BPF whose passing band contains HR. Finally, the zero-crossing technique is applied to extract the IBI estimations from the heartbeat signal.

5.5.2 Overall Performance

Fig. 5.8 shows the overall IBI estimation accuracy of the proposed mmHRV and BPFB methods. The experiment consists of 11 participants while 15 dif-

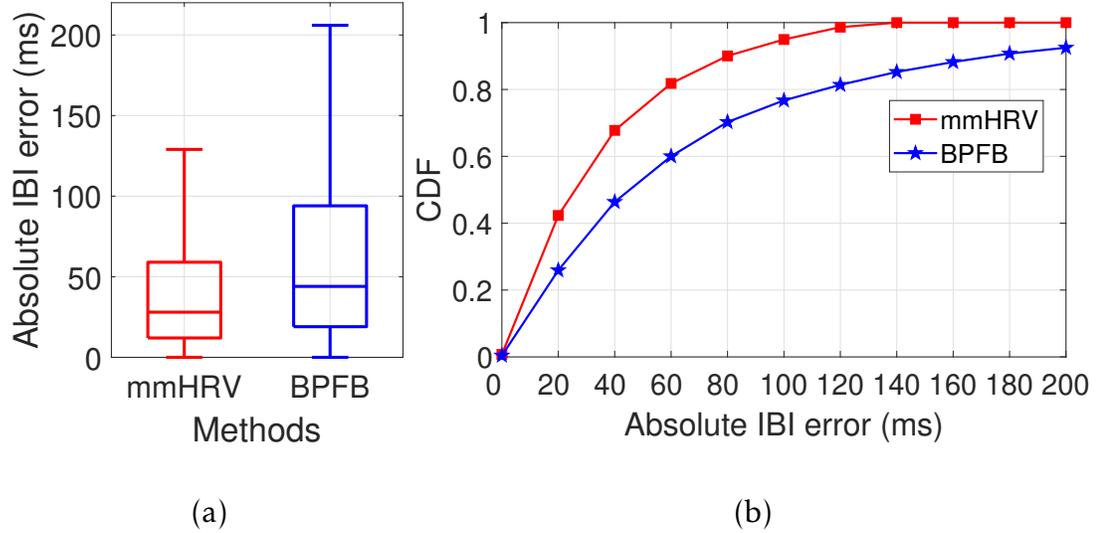


Figure 5.8: Over all performance of the IBI estimation error.

ferent experiment settings (e.g., different distances, incidental angle, orientation and blockages) are conducted for each participant. As shown in Fig. 5.8, BPFB yields about 44ms medium error while the 90-percentile error is about 180ms. The proposed mmHRV achieves a medium error of about 28ms, with the 80ms of the 90-percentile error, which outperforms the BPFB about 56%. To thoroughly evaluate the HRV estimation accuracy, Table. 5.1 and Table. 5.2 show the estimated HRV features in terms of mean IBI, RMSSD, SDRR and pNN50 of 11 participants, where the distance between user and device is about 1m. It is shown that mmHRV can achieve 3.83ms average error of mean IBI, 6.43ms average error of RMSSD, 6.45ms average error of SDRR and 2.25% average error of the pNN50. Correspondingly, the average estimation error of BPFB is 15.33ms of mean IBI, 41.94ms of RMSSD, 32.59ms of SDRR and 12.18% of the pNN50 estimations.

Table 5.1: HRV estimation results in terms of mean IBI, RMSSD, SDRR and pNN50 for 11 subjects (part-I).

Metrics	Methods		User ID					
			1	2	3	4	5	6
Mean IBI	Value (ms)	ECG	899.4	789.9	723.2	854.6	654.5	822.9
		mmHRV	906.3	790.4	725.6	848.6	652.4	828.3
		BPFB	881.5	784.2	781.5	842.1	676.6	821.7
	Error (ms)	mmHRV	6.95	0.45	2.47	5.92	2.17	5.4
		BPFB	17.87	5.7	58.36	12.44	22.01	1.25
RMSSD	Value (ms)	ECG	38.59	10.85	37.56	31.49	34.05	35.1
		mmHRV	33.52	16.53	39.08	35.26	20.29	39.72
		BPFB	59.34	54.26	53.83	52.94	78.57	65.63
	Error (ms)	mmHRV	5.08	5.68	1.52	3.77	13.76	4.62
		BPFB	20.75	43.41	16.27	21.45	44.53	30.52
SDRR	Value (ms)	ECG	56.28	22.91	50.54	35.35	33.61	48.55
		mmHRV	43.22	27.25	53.3	45.88	33.54	48.53
		BPFB	71.01	47.28	110.29	58.92	69.68	55.11
	Error (ms)	mmHRV	13.07	4.34	2.76	10.53	0.07	0.02
		BPFB	14.72	24.37	59.74	23.57	36.07	6.55
pnn50	Value (%)	ECG	11.54	0	9.15	4.32	1.14	6.29
		mmHRV	8.46	1.33	7.93	5.76	2.2	6.99
		BPFB	19.4	18.54	14.57	20	14.2	22.92
	Error (%)	mmHRV	3.08	1.33	1.22	1.44	1.05	0.7
		BPFB	7.86	18.54	5.42	15.68	13.06	16.62

Table 5.2: HRV estimation results in terms of mean IBI, RMSSD, SDRR and pNN50 for 11 subjects (part-II).

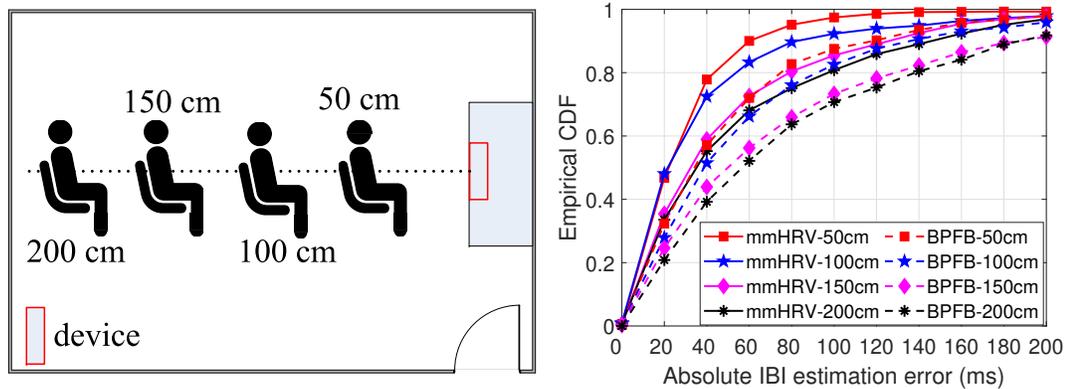
Metrics	Methods		User ID				
			7	8	9	10	11
Mean IBI	Value (ms)	ECG	645.2	890.1	564.9	728.1	763.8
		mmHRV	644.2	888.1	574.2	722.7	762.6
		BPFB	651.5	878.4	579.1	719	773.5
	Error (ms)	mmHRV	0.99	1.97	9.33	5.38	1.2
		BPFB	6.31	11.66	14.21	9.16	9.66
RMSSD	Value (ms)	ECG	16.88	27.52	5.26	23.28	31.16
		mmHRV	18.14	26.06	27.8	30.52	34.92
		BPFB	95.09	45.56	140.36	59.61	47.92
	Error (ms)	mmHRV	1.26	1.46	22.53	7.25	3.76
		BPFB	78.21	18.04	135.1	36.34	16.76
SDRR	Value (ms)	ECG	23.24	32.66	12.25	35.83	50.87
		mmHRV	25.49	37.43	38.66	37.15	45.51
		BPFB	67.61	50.44	118.41	47.92	63.94
	Error (ms)	mmHRV	2.24	4.78	26.42	1.31	5.36
		BPFB	44.37	17.78	106.16	12.09	13.07
pnn50	Value (%)	ECG	0.55	3.76	0	0.61	4.49
		mmHRV	2.17	2.26	4.83	6.71	6.41
		BPFB	18.13	12.59	10.24	12.8	12.42
	Error (%)	mmHRV	1.62	1.5	4.83	6.09	1.92
		BPFB	17.58	8.83	10.24	12.19	7.93

5.5.3 Impact of Distance

In this section, we explore the impact of the distance between the human subject and the device. As shown in Fig. 5.9 (a), the participants are asked to face towards the device and sit in four different locations ranging from 50cm to 200cm. The empirical Cumulative Distribution Function (CDF) of the absolute IBI estimation error is shown in Fig. 5.9 (b).

Fig. 5.9 (b) shows that the medium errors of mmHRV are 22ms, 22ms, 30ms and 33ms corresponding to the distance of 50cm, 100cm, 150cm and 200cm. It is clear that the IBI estimation accuracy degrades with the increment of distance, which is due to the attenuation property of the mmWave signals, as a longer propagation distance results in a lower SNR.

The medium error of BPFb increase from 40ms to 60ms with the distance increasing from 50cm to 200cm. In all the four settings, mmHRV shows better performance than the benchmark BPFb. This is because that mmHRV directly extracts the heartbeat signal from the composite signal by optimizing the decomposition, so that the error propagation from breathing as well as random body motion elimination can be avoided. Besides, the accurate heart rate estimation is necessary for BPFb method, which however is vulnerable to noise and interference from other signal components.

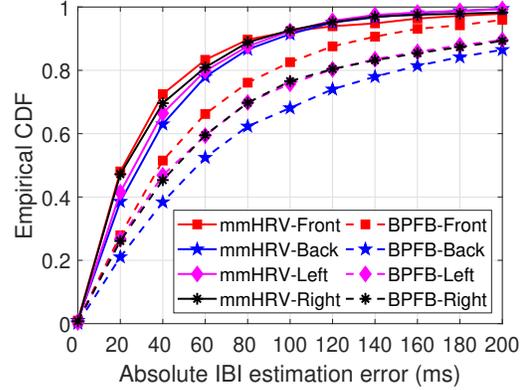
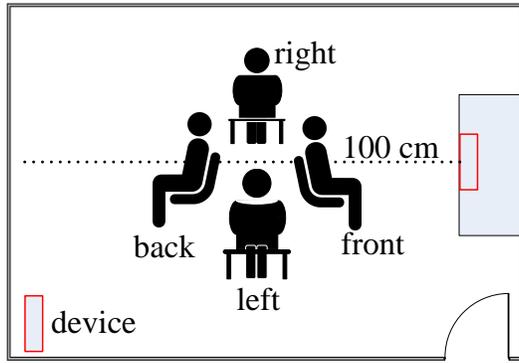


(a) Experiment setup (b) CDF of absolute IBI estimation error

Figure 5.9: Experiment setup and the absolute IBI estimation results versus distance.

5.5.4 Impact of Orientation

Considering the real case that the user may not strictly face towards the device, this section studies the impact of users' orientation on IBI estimation accuracy. As shown in Fig. 5.10 (a), four normal orientations including front, back, left and right are investigated. For all the orientations, the human subject sits 1m away from the device and Fig. 5.10 (b) shows the IBI estimation accuracy in terms of CDF. As shown in Fig. 5.10 (b), the median absolute IBI estimating error of mmHRV are 21ms, 22ms, 25ms and 28ms corresponding to the front, right, left and back settings. BPFb shows larger IBI errors with 40ms, 42ms, 43ms, 55ms correspondingly. Averagely, mmHRV outperforms BPFb of about 18.883ms in terms of RMSE. However, for both methods, the "front" setting shows the best performance while the "back" one yields the largest IBI estimation error. This is due to the physiological structure of a human body, where the vibration caused



(a) Experiment setup

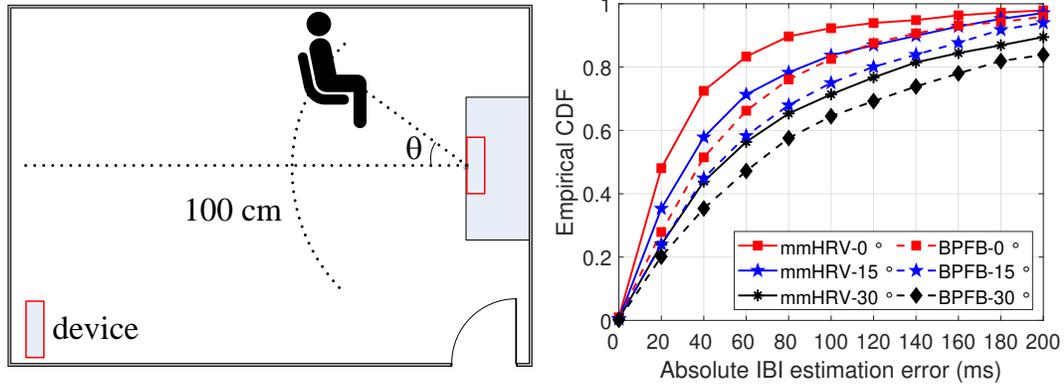
(b) CDF of absolute IBI estimation error

Figure 5.10: Experiment setup and the absolute IBI estimation results versus orientation.

by the heartbeat is larger in the front chest than in the back.

5.5.5 Impact of Incident Angle

In this section, we investigate the impact of incident angle denoted by θ in Fig. 5.11 (a). Specifically, the incident angle θ is set as $\theta \in \{0^\circ, 15^\circ, 30^\circ\}$ while the distance between the user and device is fixed at 1m. The IBI estimation errors are shown by the CDF in Fig. 5.11 (b). As expected, for both methods, the performance degrades with the increment of θ . This is because the effective reflection area decreases when the human subject deviates from the device from 0° to 30° . Moreover, according to the array signal processing theorem, the beam width will also increase with the increment of the incident angle, which reduces the directionality of the receiving signal. As a result, the SNR of the received signal decreases when the incident angle rises from 0° to 30° , thus resulting in



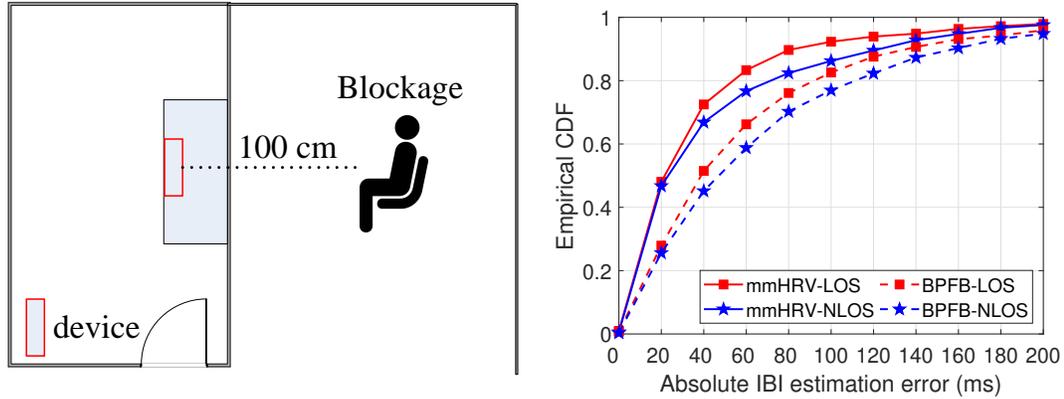
(a) Experiment setup (b) CDF of absolute IBI estimation error

Figure 5.11: Experiment setup and the absolute IBI estimation results versus incident angle.

larger IBI estimation errors. However, mmHRV still outperforms BPFb of an average about 14.544ms in RMSE, which benefits from its optimization in signal decomposition for heartbeat signal extraction as introduced in Section. 5.5.3.

5.5.6 LOS vs NLOS

This section evaluates the estimating performance when the user and the device are blocked by a wood panel as shown in Fig. 5.12 (a). The distance between the participant and the device is set as 1m while the user is asked to face towards the device. As shown in Fig. 5.12 (b), the medium estimating error of IBI of mmHRV increases from 22ms to 24ms if the blockage happens. Correspondingly, the medium error of IBI of BPFb increases from 40ms to 48ms when the blockage occurs (see Fig. 5.12 (b)). The performance degradation in the blockage setting is because that the EM signal further attenuates when it penetrates the



(a) Experiment setup

(b) CDF of absolute IBI estimation error

Figure 5.12: Experiment setup and the absolute IBI estimation results versus blockage.

wood panel, thus rendering the decrements of SNR in the received signal.

5.5.7 Impact of User Heterogeneity

To validate the robustness of mmHRV over different users, Fig. 5.13 summarizes the absolute IBI estimation error distribution for all the 11 users of different settings (including different distance, incident angles, orientations and blockage scenario). Fig. 5.13 shows the error distribution of each user, where the first 6 users are males denoted by the blue box and the last 5 users are females denoted by the red box. Evidently, mmHRV demonstrates different IBI estimation errors for different users in which the medium error varies from 13.5ms to 37ms. This can be caused by several reasons such as different body shapes and heartbeat strengths over different users. It is shown that the 75-percentile error of all the uses are smaller than 75ms, which indicates great robustness of mmHRV over

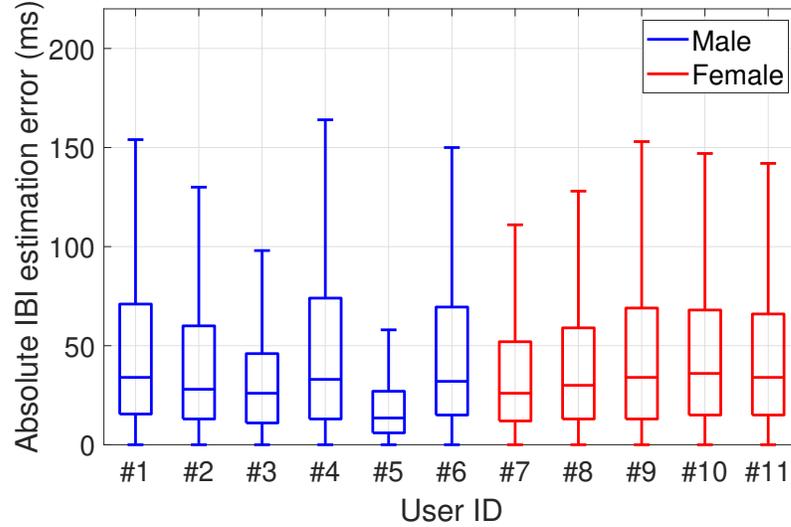
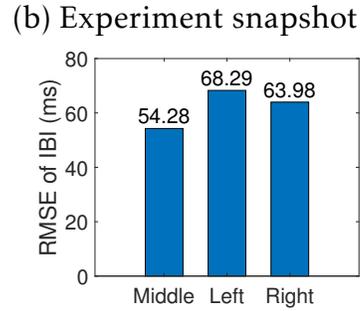
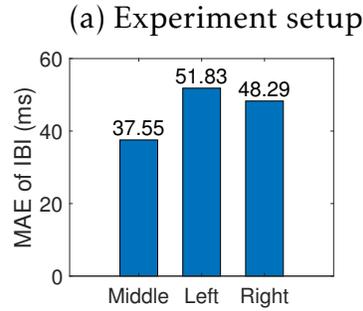
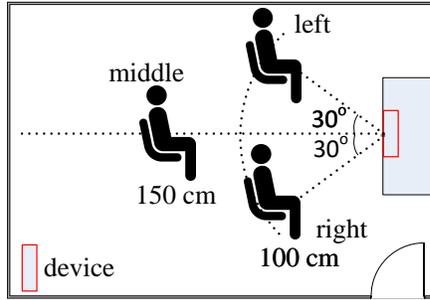


Figure 5.13: Impact of user heterogeneity.

different subjects.

5.5.8 Multiple-User Case

In this section, we investigate the accuracy of mmHRV in a multiple-user scenario. As shown in Fig. 5.14 (a) and Fig. 5.14 (b), the participant in the middle is 1.5 meters away from the device with incident angle 0° , while the other 2 users in the left and right are 1m away from the device at incidental angle $\pm 30^\circ$. Fig. 5.14 (c) and Fig. 5.14 (d) depicts the MAE and RMSE of the IBI estimation for mmHRV. The target detection result is shown in Fig. 5.4 (d). Overall, mmHRV can work robustly for the 3-user setup, where the MAE of the IBI estimation is less than 51.83ms for all the 3 locations, as shown in Fig. 5.14 (c). The RMSE of the IBI estimation is within 70ms for all the locations. We can see that mmHRV achieve higher accuracy in the middle location than that of either the left or the



(c) MAE of IBI estimation.

(d) RMSE of IBI estimation.

Figure 5.14: Mean and RMSE of the absolute IBI estimation error of multiple users.

right location. This is because that the participant at the middle location enjoys the larger reflection area and thus achieves higher SNR in the received signal correspondingly. This result coincides with the experiment result as shown in Section. 5.5.3 and Section. 5.5.5. However, the accuracy decreases compared with the single-user scenario for all the 3 locations. The main reason is that mmHRV utilizes the digital beamforming, and thus the reflections from other people, although suppressed by digital beamforming, act as extra interference compared with the single-user case.

5.6 Summary

In this chapter, we devise mmHRV, a contact-free multi-user HRV estimating system built upon a commercial mmWave radio. To identify the number of users and their locations, a target detector is first designed to locate each user without any prior calibration. The heartbeat wave of each user is then estimated by optimizing the decomposition of the composite phase information consisting respiration, heartbeats and random body motion. The exact time of heartbeats is extracted from the estimated heartbeat wave to further evaluate the IBIs and HRV metrics. Extensive experiments are conducted, where 11 participants aging from 20 to 60 are asked to sit at different locations (distance, incidental angle, orientation, and NLOS scenario) for HRV evaluation. Experimental results show that mmHRV achieves a median error of 28 ms for the IBI estimation, outperforming the state-of-art work.

Chapter 6: Driver Vital Signs Monitoring

6.1 Introduction

Automobiles have become a daily necessity in current fast-paced world due to its mobility, convenience and comfortableness. Statistics show that the number of worldwide automobiles on-the-road has reached 1.2 billion by 2015 [81]. However, in the meanwhile, road traffic crashes result in the deaths of about 1.35 million people around the world each year and leave between 20 and 50 million people with non-fatal injuries [61], according to World Health Organization.

To reduce the number of road accidents and enhance the driving safety, automobile manufacturers and researchers have been working on more and more Advanced Driver Assistance Systems (ADAS). Among many popular topics in autonomous driving, driver's vital sign monitoring is one of the essential components. Continuously monitoring driver's status makes it possible to allow the ADAS to take control of the automobiles in case of emergency, such as that the driver encounters a sudden heart attack, stroke or fatigue, which can be predicted/indicated by using the driver's Heart Rate Variability (HRV), i.e., the variation of the inter-beat intervals (IBI). HRV, in combination with Heart Rate (HR) [28] and Respiration Rate (RR) [91], has been well established as a good

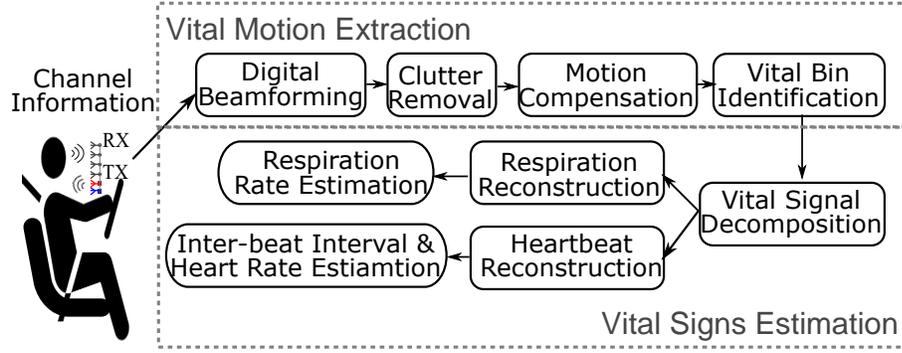


Figure 6.1: Processing flow of system.

indicator of cardiac arrhythmia, alcohol usage [4], mental stress [23] and drowsiness [18] and thus predicts the human alertness well.

In this chapter, we propose a novel system, which can estimate driver’s RR, HR and IBI considering the presence of driver’s motion artifacts using commercial mmWave radio. The rest of the chapter is organized as follows. The system is overviewed in Section 6.2, followed by vital motion extraction in Section 6.3 and vital signs estimation in Section 6.4. Section 6.5 evaluates the performance of the system. The whole chapter is summarized in Section 6.6.

6.2 System Overview

The proposed system aims at non-contact driver’s vital sign monitoring in practical driving scenarios with inevitable random motions by using a single commodity Frequency-Modulated Continuous Wave (FMCW) radar. The pipeline of the system is shown in Fig. 6.1, which consists of two main modules: **(1) Vital motion extraction** and **(2) Vital signs estimation**.

In the first stage, the *vital motion extraction* module extracts the bins con-

taining vital signals from the channel information. To begin with, conventional beamforming is performed on the channel information to get the Channel Impulse Response (CIR) at different range-azimuth bins. Then the clutter removal is performed to subtract the background reflections. However, vital signals are not directly achievable even after background subtraction because the driver's location w.r.t. radar can change over time (e.g., body roaming due to acceleration or break) during driving. As a result, the vital signals will spread over multiple range bins. Therefore, a motion compensation algorithm is devised to eliminate the effect of large body movement. The location change of the driver is first roughly compensated between consecutive CIRs based on correlation of the CIR amplitude. Then, the subtle motion within the range bin are estimated and eliminated from the CIR phase utilizing smoothing spline. After motion compensation, the range-azimuth bins containing vital signals will show periodic pattern, and the CIR of these bins will be exported for further vital signs estimation.

In the second stage, the *vital signs estimation* module estimates drivers' RR, HR and heart HRV using the vital signals exported by the previous module. To enable HRV analysis, heartbeat wave needs to be reconstructed to get the exact time of each heartbeat. However, it is non-trivial to extract the heartbeat signal from the compound vital signals including both respiration and heartbeat movements. To accurately recover the heartbeat signal, we optimize the decomposition of vital signals in all candidate bins with multiple band-limited signals concurrently. And the extracted heartbeat signals in all the candidate bins are further combined to give an estimate of the heartbeat wave. The IBIs are further

extracted from the reconstructed heartbeat wave to estimate HRV.

6.3 Vital Motion Extraction

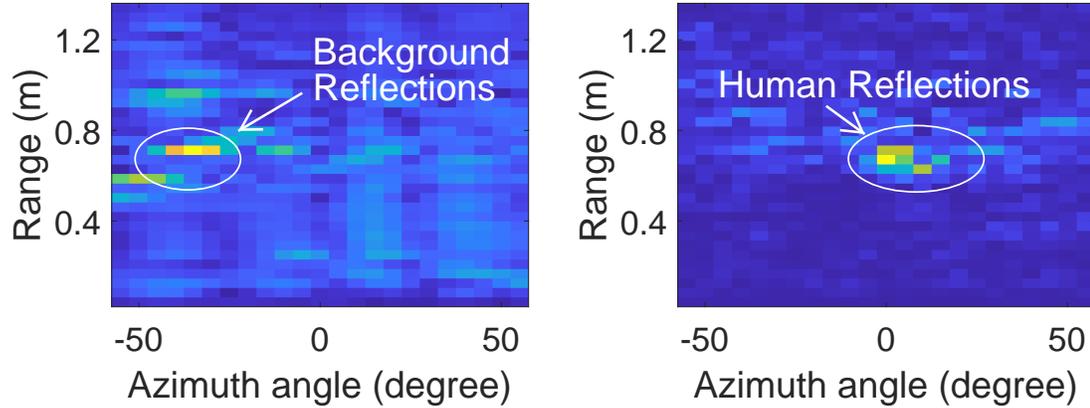
In a real-world setting, extracting vital motions from the RF signal is not trivial. Due to the presence of various clutters in car (e.g., chairs, metal objects, ceilings, etc.), it is hard to filter the RF reflections off human body. Moreover, since body motion will be involved during driving, the periodicity of reflected signal caused by vital motions can be corrupted, complicating the detection of vital signals.

6.3.1 Clutter Removal

To locate the range-azimuth bins corresponding to the driver and reduce the impact of reflections from static objects in the vehicle, the system deploys a clutter removal algorithm to subtract the CIR from the background. Note that the reflections from the static object is reasonably assumed to be invariant within a certain period of time, while the reflections from the driver change over time due to human motion (including body motion and motion caused by vital signals). The background profile can be estimated by taking average of the CIR over slow-time, and the calibrated CIR can be denoted as

$$\hat{h}(r, \theta, m) = h(r, \theta, m) - \frac{1}{M} \sum_{i=1}^M h(r, \theta, m - i), \quad (6.1)$$

where r , θ and m denote the range tap, azimuth angle and time index respectively. M is the total number of samples used for clutter removal. Fig. 6.2 shows



(a) CIR amplitude before clutter removal (b) CIR amplitude after clutter removal

Figure 6.2: Example of clutter removal.

the effect of the background cancellation, where the raw CIR before clutter removal is shown in Fig. 6.2 (a), and the corresponding calibrated CIR after clutter removal is shown in Fig. 6.2 (b). As can be seen, clutter removal reduces the background noise significantly.

6.3.2 Motion Compensation

After extracting the dynamic CIR corresponding to the driver, we would like to get the range-azimuth bins contributed by the vital signals. The vital bins can be easily identified by checking the periodicity of the phase signal if the human subject stays stationary as studied in previous works [75]. However, the assumption of the stationary human subject barely holds in the driving scenario. To recover the periodic vital signals from the CIR involving human motion, we design a two-step motion compensation algorithm, where the large body motion is compensated based on the cross correlation between consecutive CIR sam-

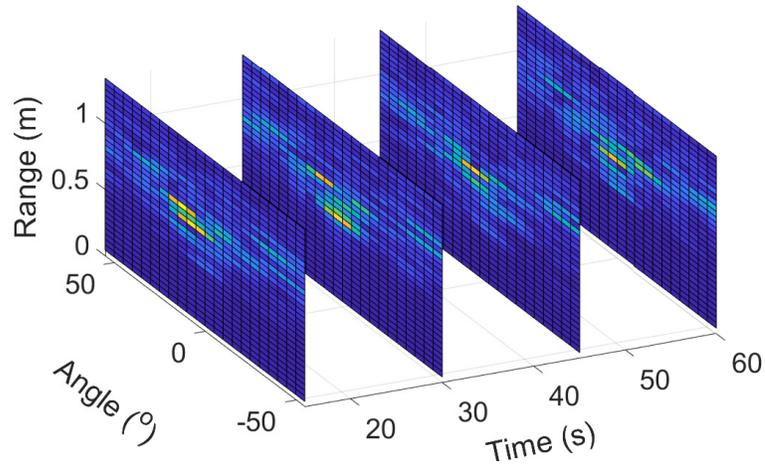


Figure 6.3: Example of consecutive frame after clutter removal.

ples, and the fine movement revealed in phase information is further removed by smoothing splines.

6.3.2.1 Large Body Movement Compensation

Note that when there is large body motion, the location of range-azimuth bins corresponding to human subject will change, as shown in Fig. 6.3, where the human subject sit at round 0.5m away from device at azimuth angle 0° . The human subject sways the body back-and-forth, resulting in the change of reflecting locations. However, the profile of human reflections stays similar, as shown in Fig. 6.3. Therefore, to remove body movement, the 2-dimensional cross correlation [9] between consecutive CIRs is calculated. Then the CIR at each time instance is circularly shifted to the point corresponding to the maximum cross correlation.

Fig. 6.4 shows the amplitude of 1-minute CIR before and after body move-

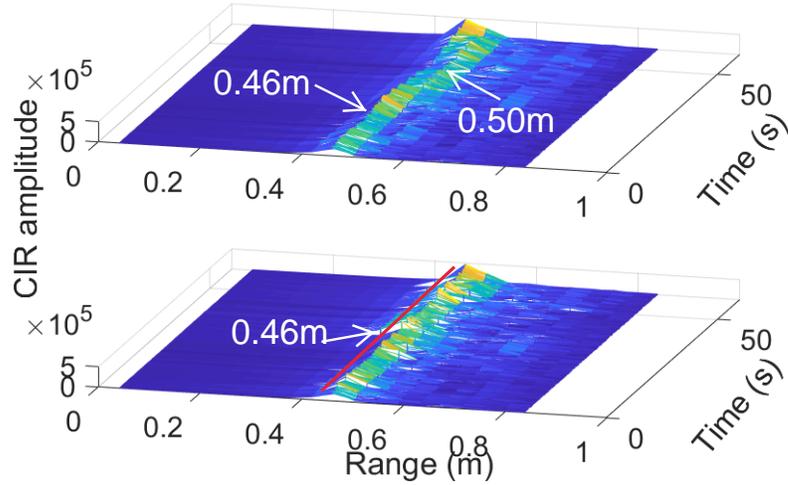
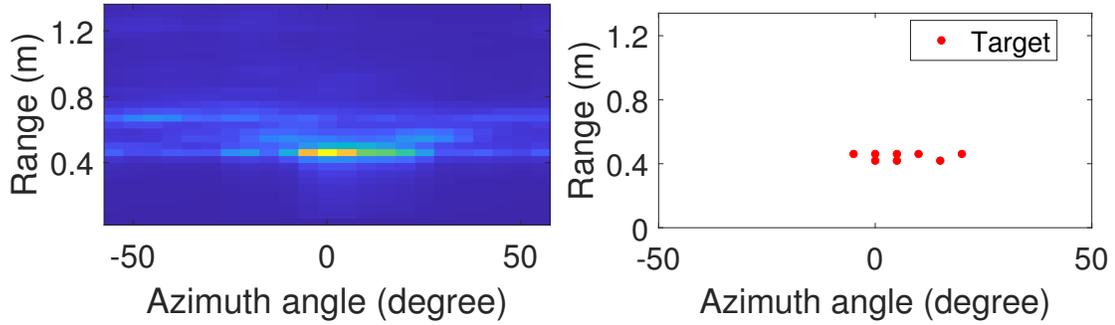


Figure 6.4: Example of large body movement compensation.

ment compensation. For visualization, we plot the CIR at azimuth angle 0° over range $[0, 0.9]$ m. It is shown that after the large body movement compensation, the bins corresponding to human subject have been aligned, as shown in the lower subfigure. The 2-D Constant False Alarm Rate (CFAR) detector will be further applied over the CIR after aligning the human subject, and the candidate bins with human subject can be selected as shown in Fig. 6.5.

Although the candidate range-azimuth bins corresponding to human subject have been aligned and selected in the first step, it is still hard to locate those bins reflected by chest with periodic vital signals. The reason is that the first step can only remove the motion artifacts that is larger than the range-azimuth resolution, however, it cannot deal with the fine movements within the range-azimuth resolution. Fig. 6.6 (a) shows an example of the unwrapped phase measurement in solid lines, and slow trend is caused by the fine movements. Therefore, to recover the periodicity of vital signals, we need to further cancel the impact of



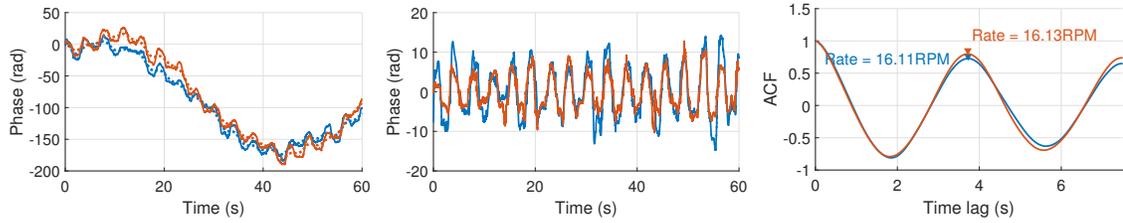
(a) CIR amplitude average over 1-min window after large body motion compensation (b) Bins corresponding to the driver by using CFAR detector

Figure 6.5: Example of target detection.

these fine movements.

6.3.2.2 Fine Movement Cancellation

Compare to the vital signs, the motion artifacts have larger distance change and lower frequency, thus, the estimation of the phase change caused by motion can be obtained by the smoothing spline algorithm, as discussed in (4.11). Fig. 6.6 illustrates the effect of fine movement cancellation, where dashed lines in Fig. 6.6 (a) show the estimated phase measurement caused by body movement. Fig. 6.6 (b) shows the phase measurement after we remove the motion artifacts, where the periodicity caused by vital signals appears. The above fine movement cancellation is performed over all candidate bins selected by CFAR detector and the cleaned phase of each candidate bin is saved for further analysis.



(a) Original unwrapped phase measurement. (b) Phase measurement after fine movement cancellation. (c) ACF of the calibrated phase measurement.

Figure 6.6: Example of fine movement cancellation.

6.3.3 Vital Bin Identification

Note that after motion compensation in Section 6.3.2, the phase information corresponding to the human chest show periodicity due to the modulation of both respiration and heartbeat, as shown in Fig. 6.6 (b). To filter out the bins reflected by other parts of human body, we check the periodicity of the phase signals over slow time by examine their Autocorrelation Function (ACF).

For the phase measurement contains vital signals, a peak can be observed at τ^* in its corresponding ACF, which reveals the time duration of a breathing cycle. Fig. 6.6 (c) shows an example of the ACF of the phase measurement corresponding to human chest, where the time duration of a breathing cycle is about 3.7s, correspond to 16.1 Respiration Per Minute (RPM). We check the periodicity over all candidate bins corresponding to the human subject, and those bins whose peak located within the range of normal human RR are identified as vital bins for further analysis.

6.4 Vital Signs Estimation

The vital bins identified by the previous module can only reflect the compound distance change caused by respiration and heartbeat. To further estimate the vital signs including RR, HR, and HRV, we need to reconstruct the distance change caused by respiration and heartbeat respectively. For simplicity, in the following analysis, we directly use the analog form of the signal model.

Let $\mathbf{y}(t) = [s_1(t), s_2(t), \dots, s_B(t)]^T$ denote the vector of the phase signals of all the B vital bins. Recall that the phase signal after movement cancellation is a mixture of vital signals, we have

$$\mathbf{y}(t) = \mathbf{s}_r(t) + \mathbf{s}_h(t) + \mathbf{n}(t). \quad (6.2)$$

where $\mathbf{s}_r(t)$ and $\mathbf{s}_h(t)$ denote the vector of respiration and heartbeat signal respectively, $\mathbf{n}(t)$ is the random phase offset introduced by noise, which is independent with the phase change caused by vital signs. To decompose the phase and get the estimate of vital signs, we leverage the following properties. First, both $\mathbf{s}_r(t)$ and $\mathbf{s}_h(t)$ are quasi-periodic signals, whose periodicity changes slightly over time. Second, the periodicity of signals stays the same in different vital bins. Third, due to the physiological structure of human body, the distance change caused by respiration and heartbeat can be different in different parts of human body (i.e., the distance change in different vital bins can be distinct).

The phase signal, therefore, can be decomposed as an ensemble of band-limited signals, denoted as $\{\mathbf{u}_k(t)\}_{k=1}^K$, where for each component $\mathbf{u}_k(t)$, which is

defined as $\mathbf{u}_k(t) = [u_{\{k,1\}}(t), u_{\{k,2\}}(t), \dots, u_{\{k,B\}}(t)]^T$, the decomposed signals w.r.t. all vital bins should be compact around the same center frequency ω_k . The decomposition is modeled as [72]

$$\begin{aligned} \min_{u_{k,b} \in \mathcal{U}, \omega_k \in \Omega} \alpha \sum_{k=1}^K \sum_{b=1}^B \left\| \partial t \left[(\delta(t) + \frac{j}{\pi t}) * u_{k,b}(t) \right] \exp(-j\omega_k t) \right\|_2^2 \\ + \sum_{b=1}^B \left\| y_b(t) - \sum_{k=1}^K u_{k,b}(t) \right\|_2^2, \end{aligned} \quad (6.3)$$

where $\mathcal{U} = \{u_{1,1}, u_{1,2}, \dots, u_{1,B}, \dots, u_{K,B}\}$ and $\Omega = \{\omega_1, \dots, \omega_K\}$ denote the set for all components and their center frequencies, respectively. The first term in (6.3) represents the bandwidth constraint, which is measured by the sum of the L_2 norm of the gradient of the analytic signal corresponding to each component. The second term is the fidelity constraint, which is evaluated by the quadratic penalty w.r.t. reconstruction. α is a parameter for balancing the bandwidth constraint and data fidelity. The optimization problem in (6.3) can be solved by alternatively updating \mathcal{U} and Ω until convergence.

6.4.1 Minimization w.r.t. $u_{k,b}$

To update the k -th component for vital bin b , the subproblem can be written as

$$\begin{aligned} u_{k,b}(t) = \arg \min_{u_{k,b}(t)} \left\| \partial t \left[(\delta(t) + \frac{j}{\pi t}) * u_{k,b}(t) \right] \exp(-j\omega_k t) \right\|_2^2 \\ + \left\| y_b(t) - \sum_{i=1}^K u_{i,b}(t) \right\|_2^2. \end{aligned} \quad (6.4)$$

By using the Parseval theorem, the problem is equivalent to

$$\begin{aligned} \Gamma_{k,b}(\omega) = \arg \min_{\Gamma_{k,b}(\omega)} \alpha & \left\| j\omega[(1 + \text{sgn}(\omega + \omega_k))\Gamma_{k,b}(\omega)] \right\|_2^2 \\ & + \left\| \dagger_b(\omega) - \sum_{i=1}^K \Gamma_{i,b}(\omega) \right\|_2^2, \end{aligned} \quad (6.5)$$

where $\Gamma_{k,b}(\omega)$ and $\dagger_b(\omega)$ are the Fourier transfer of $u_{k,b}(t)$ and $y_b(t)$ respectively.

Now, we take integrals over frequencies and we have

$$\begin{aligned} \Gamma_{k,b}(\omega) = \arg \min_{\Gamma_{k,b}(\omega)} \alpha & \int_{-\infty}^{\infty} \left\| j\omega[(1 + \text{sgn}(\omega + \omega_k))\Gamma_{k,b}(\omega)] \right\|_2^2 \\ & + \left\| \dagger_b(\omega) - \sum_{i=1}^K \Gamma_{i,b}(\omega) \right\|_2^2 d\omega. \end{aligned} \quad (6.6)$$

After performing a change of variables $\omega \leftarrow \omega - \omega_k$ in the first term, and using the Hermitian symmetry of the real signals in the spectrum for the second term, the above problem can be rewritten as

$$\Gamma_{k,b}(\omega) = \arg \min_{\Gamma_{k,b}(\omega)} \int_0^{\infty} 4\alpha(\omega - \omega_k)^2 |\Gamma_{k,b}(\omega)|^2 + 2|\dagger_b(\omega) - \sum_{i=1}^K \Gamma_{i,b}(\omega)|^2 d\omega. \quad (6.7)$$

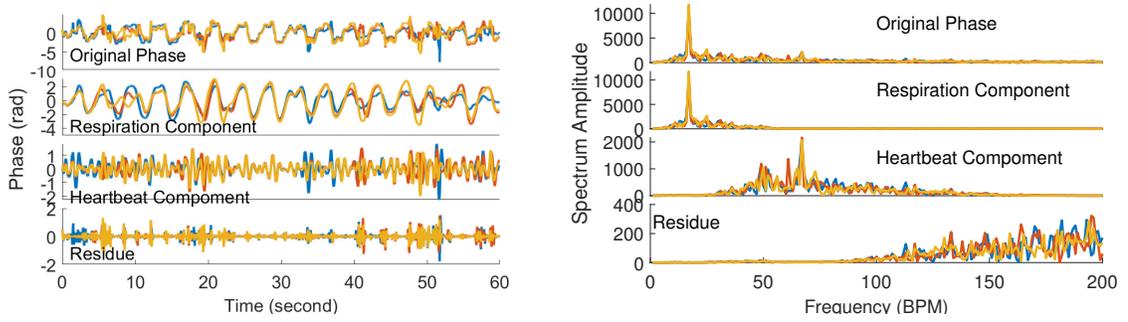
The updated solution can be expressed as

$$\Gamma_{k,b}(\omega) = \frac{\dagger_b(\omega) - \sum_{i,i \neq k} \Gamma_{i,b}(\omega)}{1 + 2\alpha(\omega - \omega_k)^2}. \quad (6.8)$$

6.4.2 Minimization w.r.t. ω_k

The center frequencies ω_k only appears in the bandwidth constraint and thus the updating function can be written as

$$\omega_k = \arg \min_{\omega_k} \sum_{b=1}^B \left\| \partial t \left[(\delta(t) + \frac{j}{\pi t}) * u_{k,b}(t) \right] \exp(-j\omega_k t) \right\|_2^2. \quad (6.9)$$



(a) Decomposition of a typical phase signal in time domain. (b) Corresponding spectrum of the decomposed components.

Figure 6.7: Example of phase decomposition of 3 vital bins.

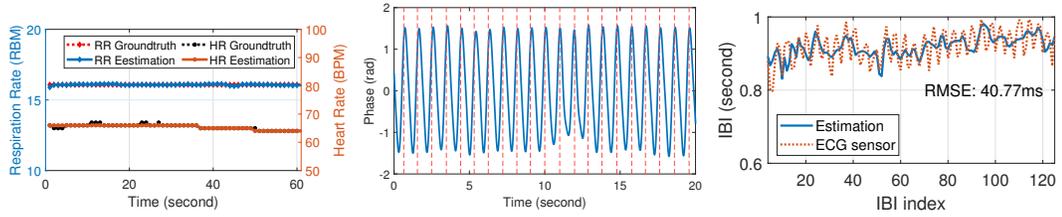
As before, we find the optimum in Fourier domain, and we have

$$\omega_k = \arg \min_{\omega_k} \sum_{b=1}^B \int_0^{\infty} (\omega - \omega_k)^2 |\Gamma_{k,b}(\omega)|^2 d\omega. \quad (6.10)$$

The minimizer of the above quadratic problem is

$$\omega_k = \frac{\sum_b \int_0^{\infty} \omega |\Gamma_{k,b}(\omega)|^2 d\omega}{\sum_b \int_0^{\infty} |\Gamma_{k,b}(\omega)|^2 d\omega}. \quad (6.11)$$

Fig. 6.7 shows an example of vital signals decomposition, where the time and frequency domain of the original phase as well as the decomposition components are shown in Fig. 6.7 (a) and Fig. 6.7 (b) respectively. The information of 3 different vital bins are distinguished by the color of lines, and it is clear to see that although the distance change of different vital bins are distinct, as shown in Fig. 6.7 (a), the periodicity of the signal of each component stays the same, as shown in Fig. 6.7 (b). In other words, components corresponding to vital signals are perfectly aligned over all vital bins, e.g., the first component represents the distinct displacement cause by respiration over different vital bins.



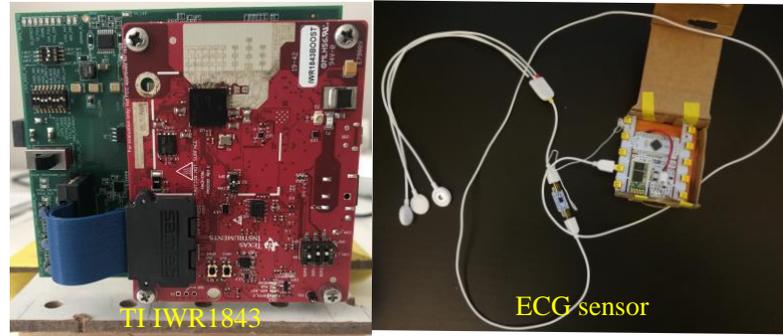
(a) RR and HR estimation. (b) Example of estimated heartbeat wave. (c) IBI estimation.

Figure 6.8: Example of estimated result v.s. ground truth.

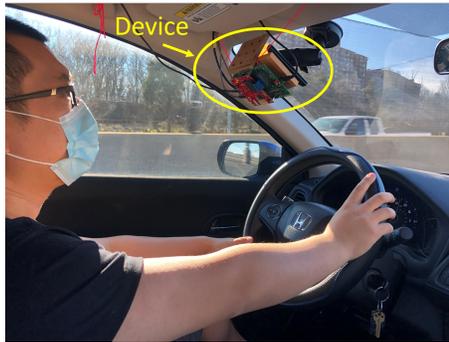
6.4.3 Vital Signals Reconstruction

To further reduce the noise impact, we reconstruct the vital signals by combining the signals of all vital bins using empirical mean, i.e., $s_r(t) = \frac{1}{B} \sum_b u_{i,b}(t)$ and $s_h(t) = \frac{1}{B} \sum_b u_{j,b}(t)$, where i -th and j -th components correspond to respiration signal and heartbeat signal respectively. The RR is estimated by finding the first peak of the ACF of the estimated respiration signal, as shown in Fig. 6.6 (c). Besides, the FFT is further performed on the estimated heartbeat signal to get the estimation of HR. The exact time of each heartbeat can be further extracted from the reconstructed heartbeat wave to estimate IBI.

Fig. 6.8 shows the estimated vital signs versus their ground-truths of a 2-minute dataset, where a 1-minute window is employed for the time-frequency domain transform (i.e., ACF and FFT). The estimated RR and HR are shown in solid lines in Fig. 6.8 (a), which match with the ground-truth, shown as dashed lines in Fig. 6.8 (a). Fig. 6.8 (b) shows a segment of estimated heartbeat wave, and the ground-truth of the exact time of each heartbeat is marked as vertical dashed



(a) Hardware.



(b) Device mount on windshield.



(c) Device under steering wheel.

Figure 6.9: Experiment setup.

lines. The estimated IBIs of the whole data and their corresponding ground-truth are shown in Fig. 6.8 (c). Evidently, the proposed system achieves high accuracy in vital signs estimation, and the Root-Mean-Squared-Error (RMSE) of IBI estimation in Fig. 6.8 (c) is 40.77ms, corresponding to the 96% relative accuracy.

6.5 Experiment Evaluation

In this section, extensive experiments are performed to evaluate the performance of the proposed system. We compare the performance of the proposed system with the state-of-art work under different experimental settings.

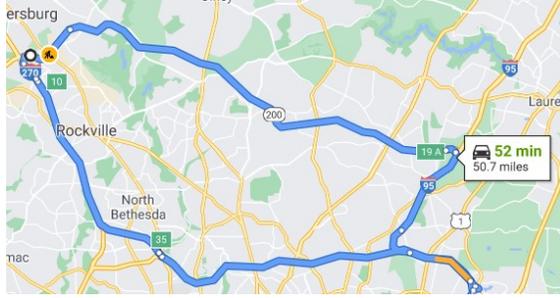


Figure 6.10: Experiment path.

6.5.1 Methodology

We conduct experiments using a Commodity Off-The-Shelf (COTS) mmWave radar, IWR1843BOOST [26], as shown in Fig. 6.9 (a), where the 2 Tx antennas and 4 Rx antennas are configured in TDM-MIMO mode [46]. The device can achieve a theoretical azimuth resolution of 15° , and the field-of-view (FoV) is 100° in horizontal plane, which is large enough to cover the driver. The ground truth of heartbeat is captured by a commercial ECG sensor [25], as shown in Fig. 6.9 (a), and the ground truth of breathing is measured by a respiration belt [19].

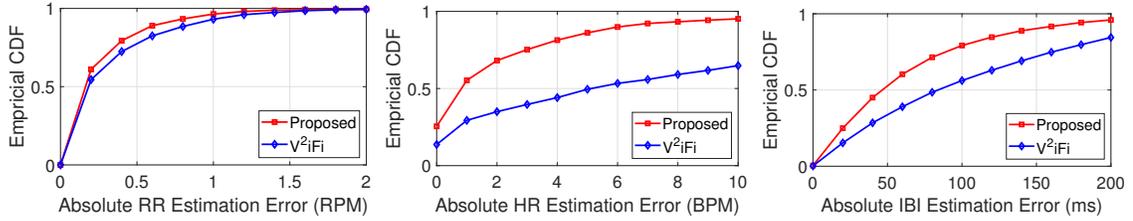
We recruit 4 volunteers (2 males and 2 females) to help on the data collection including 2 different device locations as shown in Fig. 6.9 (b) and Fig. 6.9 (c). The driving route is a cycle of 50.7 miles including local routes and highway, where the road conditions can be referred to Maryland’s GIS DataSet [42]. During the data collection, the driver is driving following their own habits with no further constraints, and a copilot is responsible for collecting data.

To further evaluate the performance of the proposed system, we compare it with the state-of-art work, V²iFi [94], which also estimates driver’s vital signs

using mmWave FMCW radar. With assumption that the distance change caused by vital signals are identical in different vital bins, V²iFi estimates the respiration and heartbeat signal by Multi-Sequence Variational Mode Decomposition (MS-VMD). Note that V²iFi cannot estimate vital signs when drivers have body motion. For fair comparison, motion compensation proposed in Section 6.3.2 is also applied to V²iFi to remove the motion artifacts before estimating the vital signals.

6.5.2 Overall Performance

Fig. 6.11 depicts the overall performance of the proposed system and V²iFi. The experiments consist of road tests with different pavement conditions, device locations, as well as the controlled experiments with different motion types, including stationary, head motion, hand motion and back-and-forth torso motion for 4 different users. Fig. 6.11 (a) plots the empirical cumulative CDF of absolute RR estimation error, where the 90-percentile error for the proposed system and V²iFi are 0.64 RPM and 0.86 RPM respectively. The performance improvement is more significant for HR estimation, where the proposed system achieves a median error of 0.82 BPM, and the median error of V²iFi is 5.12 BPM, as shown in Fig. 6.11 (b). Fig. 6.11 (c) shows the performance of IBI estimation for the two systems, where V²iFi yields about 84ms medium error, while the proposed system achieves a medium error of 46ms, outperforming V²iFi about 45.2%.



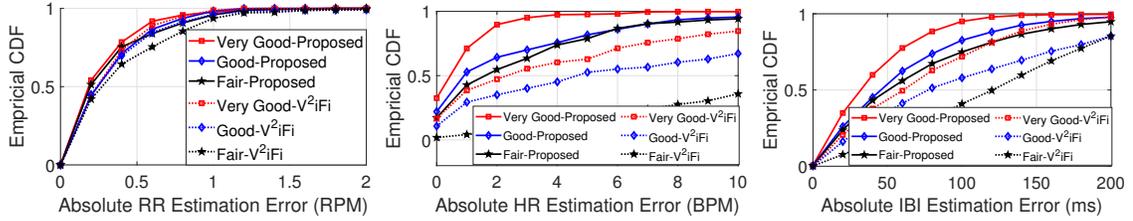
(a) CDF of RR estimation error. (b) CDF of HR estimation error. (c) CDF of IBI estimation error.

Figure 6.11: Comparison of vital sign estimation performance between proposed method and V²iFi.

6.5.3 Impact of Road Condition

In this section, we investigate the effect of road condition on the estimation accuracy. Note that the road/pavement condition is assessed by several factors including rutting, friction, structural cracking density, etc. The better the pavement condition is, the smoother the road is, and less body motion induced by uneven road will be involved. The test route shown in Fig. 6.10 roughly includes three different pavement conditions, i.e., *Very Good*, *Good* and *Fair*, and the pavement detail can be referred to Maryland’s GIS DataSet [42]. During data collection, the copilot saves data every 2 minutes and records the corresponding pavement condition at the same time.

As expected, the performance degrades with the deterioration of the pavement condition, as shown in Fig. 6.12. The medium error of RR estimation of the proposed system is 0.18RPM when the pavement condition is *Very Good*, and it increases to 0.19RPM when the pavement condition is *Fair*, as shown in Fig. 6.12 (a). The degradation is more severe in terms of HR estimation, where



(a) CDF of RR estimation error. (b) CDF of HR estimation error. (c) CDF of IBI estimation error.

Figure 6.12: Vital sign estimation performance versus pavement condition.

the median error increases from 0.45BPM to 1.64BPM when the pavement condition deteriorates from *Very Good* to *Fair*, as shown in Fig. 6.12 (b). Fig. 6.12 (c) shows the performance of IBI estimation, where the system achieves a median error of 32ms when the pavement condition is *Very Good*, and yields about 50ms of median error when the pavement condition is *Fair*. The reason for the degradation of the performance with poorer road condition is that more motion artifacts cause by the uneven road will be involved with worse pavement condition, which means a lower SINR of vital signals. Since the distance change caused by respiration is larger than heartbeat, which means a higher SINR of the RR estimation. Therefore, we observe a slighter degradation in RR estimation compared with HR and IBI estimations.

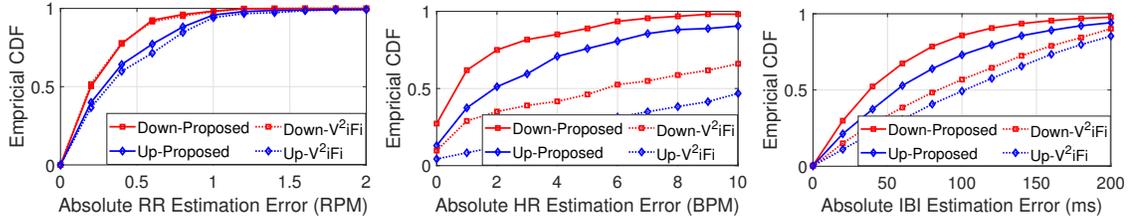
It is obvious that both algorithms achieve similar performance in RR estimation, but the proposed system outperforms V²iFi for all 3 road conditions in terms of HR and IBI estimation. In specific, for HR estimation, V²iFi yields a median error of 2.34BPM for the *Very Good* pavement condition, and the performance gap is more severe with the deterioration of road condition, as shown in Fig. 6.12 (b), where the median error for the *Good* pavement condition is

4.65BPM, which is 4 times worse than the proposed method. As for IBI estimation, V²iFi yields a median error of 61ms with the *Very Good* pavement condition, which is about 90.6% worse than the proposed method. The performance gap also becomes larger with the degeneration of the pavement condition, where for the *Fair* condition, the median error of V²iFi is 120ms, 1.4 times larger compared with the proposed system. The main reason for the performance gap between V²iFi and the proposed method is that V²iFi assumes the same distance change of vital signals in different vital bins, which is hard to meet when the SNR of the signal is small. However, the proposed system only assumes the same periodicity of vital signs in different vital bins when reconstructing vital signals, and the distance change in different vital bins are jointly optimized, which is more robust to the noise.

6.5.4 Impact of Device Location

In this study, we investigate the impact of device location on vital signs estimation. The radar is placed at the top of windshield, as shown in Fig. 6.9 (b), and under the steering wheel, as shown in Fig. 6.9 (c). Fig. 6.13 plots the CDF of the absolute error of RR, HR and IBI estimations, where the red lines corresponds to the setting when the device is placed under the steering wheel, denoting as “down” setting, and the blue lines corresponds to the setting when the device is mounted on the top of the windshield, denoting as “up” setting.

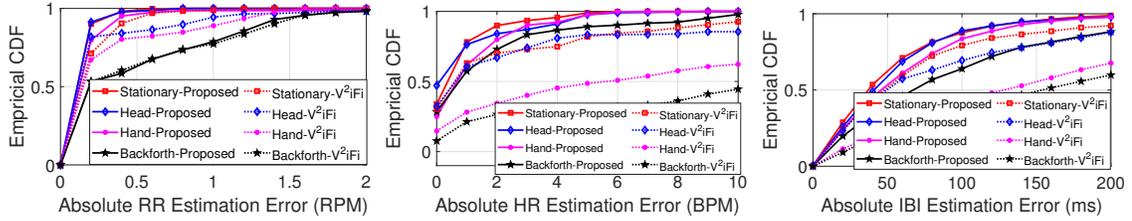
It is shown that the “down” setting achieves better performance for all es-



(a) CDF of RR estimation error. (b) CDF of HR estimation error. (c) CDF of IBI estimation error.

Figure 6.13: Vital sign estimation performance versus device location.

timations. In specific, for the proposed system, the median error for RR, HR and IBI estimation are 0.2RPM, 0.65BPM and 38ms respectively for the “down” setting, however, it increases to 0.28RPM, 1.91BPM and 56ms for the “up” setting, corresponding to 40%, 193.85% and 47.37% performance degradation, respectively. We observe the similar phenomenon in V^2iFi , where the median error for all the 3 metrics increases when the device is place as the “up” setting, as shown in dashed lines in Fig. 6.13. The reason is that when the device is mounted on the windshield, the vital bins mainly correspond to the chest, whereas, for the “down” setting, the vital bins mainly correspond to the lower chest and the abdomen. Note that for the same scenario (e.g., car decelerates due to brake), severer motion will be involved in the upper chest than the abdomen, therefore, the SNR of vital signals for the “down” setting is larger than the “up” setting. However, comparing to the proposed system, V^2iFi yields larger estimation error for all the 3 metrics, which is resulted by its less robustness to noise as discussed in Section 6.5.3.



(a) CDF of RR estimation error. (b) CDF of HR estimation error. (c) CDF of IBI estimation error.

Figure 6.14: Vital sign estimation performance versus motion type.

6.5.5 Impact of Motion Type

As driving involves different kinds of motion of head, hand and body such as looking at the side mirror, or hand motion associated with the steering operation, etc., to better understand the impact of different motion types, we conduct controlled experiment and analyze their corresponding impact in this section, as shown in Fig. 6.14. During the experiment, drivers are asked to perform specific type of motion in a parked car, including sitting stationary, head motion to check the surroundings, hand motion to operate steering wheel and randomly sway their body back-and-forth to emulate the body motion caused by acceleration and deceleration.

Fig. 6.14 (a) shows the CDF of RR estimation error with different motion types, where we can see that the median estimation error when driver performs head motion is nearly the same as the stationary case. The performance slightly degrades when the driver performs hand motion, where the median error increase from 0.11RPM to 0.12RPM compare to the stationary setting. However, for the large back-and-forth motion, we observe a severe performance degrada-

tion, and its median error of RR estimation is 0.19RPM, 72.73% worse than the stationary setting. Similar performance degradation can be observed in terms of HR and IBI estimation.

Fig. 6.14 (b) shows that the median error of HR estimation increases from 0.35BPM corresponding to the stationary setting to 0.68BPM and 0.75BPM when the driver performs hand and back-and-forth motion, respectively. As for IBI estimation, the median error when the driver performs sitting stationary, head motion, hand motion and random back-and-forth motion are 37ms, 41ms, 45ms and 68ms respectively, as shown in Fig. 6.14 (c).

We also plot the estimation performance of V²iFi in dashed lines in Fig. 6.14, where the similar performance degradation can be observed. However, we can see that V²iFi is more vulnerable to motion artifacts, and the performance degradation of hand and back-and-forth motion is more severe compared to the proposed system. In specific, we can see that the median error of HR estimation for the back-and-forth setting is larger than 10BPM, which is almost useless for driver's HR estimation.

6.5.6 Impact of User Heterogeneity

In this part, we study the impact of the user heterogeneity on the performance. Fig. 6.15 summaries the absolute IBI estimation error of 4 drivers using the data of all the settings above. Evidently, the proposed method demonstrates different IBI estimation errors for different users, where the medium error varies

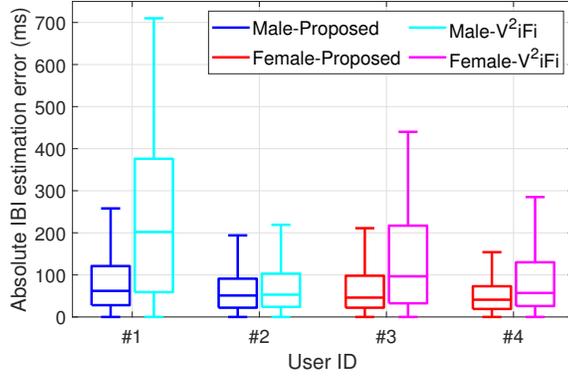


Figure 6.15: Impact of user heterogeneity.

from 41ms to 62ms, as shown in blue and red boxes in Fig. 6.15. The difference in error distribution can be caused by various factors, such as different driving habits and heartbeat strength over individuals. Besides, the performance of V²iFi is shown in cyan and magenta boxes for comparison. It is obvious that the proposed method outperforms V²iFi for all the 4 users, which is benefited from its dedicate design to resist motion artifacts.

6.6 Summary

In this chapter, we propose a novel system that can accurately detect driver’s vital signs in the presence of practical driving motions using the reflections of RF signals off the human subject only. To locate the reflections from the driver, the system first performs conventional beamforming to get the CIR with different range-azimuth bins, followed by a clutter removal module to remove the reflection from the background. Then the 2-dimensional correlation between different CIR samples have been used to eliminate large displacement caused by body

roaming. Finer motion artifacts are further removed by the smoothing spline, which can accurately estimate motion artifacts without dedicated choice of hyperparameter as in polynomial fitting. The displacement caused by respiration and heartbeat are then estimated by jointly optimizing the decomposition of vital signals in all vital bins, and the RR, HR and IBI can be extracted from the reconstructed respiration and heartbeat wave. We prototype our system using a commercial millimeter-wave radio, and conduct experiments to evaluate the performance. Experimental results show that the proposed system can estimate vital signs accurately with driving motion artifacts, outperforming the state-of-art works.

Chapter 7: Conclusions and Future Work

7.1 Conclusions

Human-centric sensing via wireless radio frequency has attracted an increasing interest for IoT applications. Since human activities can affect the propagation of wireless signals, wireless sensing has been proposed as the technique to reveal information of human subjects from ambient radio signals. In this dissertation, we demonstrate the feasibility as well as the capability of wireless sensing in vital signs monitoring by proposing one respiration tracking system using WiFi and three heart rate monitoring systems using mmWave radio. The system performance has been evaluated by extensive experiments.

In Chapter 3, we propose a system that is capable of continuously tracking the breathing rates of multiple users using the CSI of a single pair of commercial WiFi device. By leveraging both spectrum and temporal diversities, the system can match the estimated breathing rates in different time instances to different users. The breathing rate traces can be extracted even if some of them merge together for a short time period. Moreover, by utilizing the estimated breathing rate traces, the system can achieve some interesting applications such as crowd number estimation for smart home scenario. We prototype the proposed system

on COTS WiFi devices and evaluate it in both indoor and in-car environments, in which the results demonstrate promising performance.

In Chapter 4, we propose a multi-person respiration rate as well as heart rate monitoring system, ViMo, with 60GHz WiFi. By fully investigate both amplitude and phase information of the CIR measurement, the system can identify static objects, stationary human subjects and human in motion without any calibration. To have a robust estimate of heart rate for the stationary human subjects, the smoothing spline is applied to eliminate the respiration interference. We evaluate the performance of ViMo by various settings, including NLOS and motion artifacts, the most challenging scenarios for wireless vital signs monitoring. Experimental results show that ViMo monitor user's vital signs accurately, with a median error of 0.19RPM and 0.92BPM, respectively, for RR and HR estimation.

To get finer information of heartbeat, we propose a multi-user heart rate variability monitoring system, mmHRV, with a commercial FMCW mmWave radio in Chapter 5. We first develop a calibration-free target detector to identify the number of users and their locations. Then, the heartbeat signal of each user is estimated by decomposing the phase measurement modulated by the chest movement. The exact time of each heartbeat is estimated by finding the peak locations of the estimated heartbeat signal to further evaluate IBIs and HRV metrics. Extensive experiments were conducted, where 11 participants aging from 20 to 60 help to collect data under different settings including the NLOS scenario. Experimental results showed that mmHRV achieves a median error of 28ms for the IBI estimation, outperforming the state-of-art work.

In Chapter 6, we propose a driver vital sign monitoring system built upon a commercial mmWave radar. To enable driver's vital sign monitoring, we devise a two-step motion compensation module to eliminate the motion artifacts during driving. Then the respiration and heartbeat wave are reconstructed by jointly optimizing the decomposition over all range-azimuth bins containing vital signals. We evaluate the system performance in real driving environment, where the impact of pavement condition, device location as well as motion type are explored in the experiment. Experimental results illustrate promising performance of the proposed system, where the median error of RR, HR and IBI are 0.16RPM, 0.82BPM and 46ms, respectively.

7.2 Future Work

The development of IoT witnesses a proliferation of IoT devices. In the meanwhile, wireless sensing enables these wireless devices another functionality besides communication, that is to sense and understand the environment. In this dissertation, we have demonstrated the idea and the feasibility of wireless sensing for vital signs monitoring. However, there are several open problems and challenges to be explored and investigated to make the proposed systems more versatile and useful in real life scenarios.

Firstly, the proposed respiration tracking system captures the respiration rate traces from the spectrogram, where the frequency resolution depends on the window length of STFT. Therefore, if we want to differentiate two breathing

rate traces with similar frequency, longer window should be applied, which will cause extra delay. In the future, we would like to investigate the possibility of breathing signals separation in the time domain using the statistical difference between different breathing sources, which may enable the respiration sources separation with a shorter delay.

Secondly, in the second part of the thesis, three heartbeat monitoring systems have been designed using the mmWave radio. These systems either can capture the multi-person heartbeat when they are in stationary, as illustrated in Chapter 4 and Chapter 5, or capture the heartbeat signal for a single person with motion artifacts, as illustrated in Chapter 6. However, the scenario corresponding to multi-person with motion artifacts may happen from time to time. To make the proposed systems more useful in real life scenarios, we need to design the system that enables multi-person heartbeat detection with motion artifacts.

Lastly, the proposed vital signs monitoring system can be augmented with data mining module for human health status evaluation. Some medical research has validated that the human status such as emotion and mental stress can influence the vital signs. It is promising that some interesting applications can be achieved by enabling the human status prediction, such as the emotion recognition for the smart car to prevent the road accidents caused by road rage.

Bibliography

- [1] *Cardinal Spline Interpolation, Chapter 2, The Basis Property of B-Splines.*
- [2] IEEE P802. 11ad. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancement for Very High Throughput in the 60 GHz Band. 2012.
- [3] H. Abdelnasser, K. Harras, and M. Youssef. UbiBreathe: A Ubiquitous non-Invasive WiFi-based Breathing Estimator. In *ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015.
- [4] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri. Heart Rate Variability: A Review. *Medical and biological engineering and computing*, 2006.
- [5] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller. Smart Homes that Monitor Breathing and Heart Rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.
- [6] A. Ahmad, J. C. Roh, D. Wang, and A. Dubey. Vital Signs Monitoring of Multiple People using A FMCW Millimeter-wave Sensor. In *2018 IEEE Radar Conference (RadarConf18)*, 2018.
- [7] M. N. Alam Nipu, S. Talukder, M. S. Islam, and A. Chakrabarty. Human Identification using WIFI Signal. In *2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, 2018.
- [8] S. Bakhtiari, T. W. Elmer, N. M. Cox, N. Gopalsami, A. C. Raptis, S. Liao, I. Mikhelson, and A. V. Sahakian. Compact Millimeter-Wave Sensor for Remote Monitoring of Vital Signs. *IEEE Transactions on Instrumentation and Measurement*, 2012.
- [9] L. Brown. A Survey of Image Registration Techniques. *ACM Computing Surveys*, 1992.

- [10] J. A. Chalmers, D. S. Quintana, M. J.-A. Abbott, and A. H. Kemp. Anxiety Disorders Are Associated with Reduced Heart Rate Variability: A Meta-analysis. *Frontiers in psychiatry*, 2014.
- [11] C. Chen, Y. Han, Y. Chen, H. Lai, F. Zhang, B. Wang, and K. J. R. Liu. TR-BREATH: Time-Reversal Breathing Rate Estimation and Detection. *IEEE Transactions on Biomedical Engineering*, 2018.
- [12] C. Chen, Y. Han, Y. Chen, and K. J. R. Liu. Multi-person Breathing Rate Estimation using Time-reversal on WiFi Platforms. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016.
- [13] E. Cianca, M. De Sanctis, and S. Di Domenico. Radios as Sensors. *IEEE Internet of Things Journal*, 2017.
- [14] S. Depatla and Y. Mostofi. Crowd Counting Through Walls using WiFi. In *IEEE International Conference on Pervasive Computing and Communications*, 2018.
- [15] S. Depatla, A. Muralidharan, and Y. Mostofi. Occupancy Estimation using Only WiFi Power Measurements. *IEEE Journal on Selected Areas in Communications*, 2015.
- [16] K. Dragomiretskiy and D. Zosso. Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 2014.
- [17] J. Morales et al. Use of Heart Rate Variability in Monitoring Stress and Recovery in Judo Athletes. *The Journal of Strength & Conditioning Research*, 2014.
- [18] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda, et al. Heart Rate Variability-based Driver Drowsiness Detection and Its Validation with EEG. *IEEE Transactions on Biomedical Engineering*, 2018.
- [19] Go Direct Respiration Belt. <https://www.vernier.com/product/go-direct-respiration-belt/>.
- [20] A. De Groote, M. Wantier, G. Cheron, M. Estenne, and M. Paiva. Chest Wall Motion During Tidal Breathing. *Journal of Applied Physiology*, 1997.
- [21] C. Gu, G. Wang, T. Inoue, and C. Li. Doppler Radar Vital Sign Detection with Random Body Movement Cancellation Based on Adaptive Phase Compensation. In *IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013.
- [22] C. Gu, G. Wang, Y. Li, T. Inoue, and C. Li. A Hybrid Radar-Camera Sensing System with Phase Compensation for Random Body Movement Cancellation in Doppler Vital Sign Detection. *IEEE Transactions on Microwave Theory and Techniques*, 2013.

- [23] Y. Han, T. Lauteslager, T. S. Lande, and T. G. Constandinou. UWB Radar for Non-contact Heart Rate Variability Monitoring and Mental State Classification. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019.
- [24] W. Hu, Z. Zhao, Y. Wang, H. Zhang, and F. Lin. Noncontact Accurate Measurement of Cardiopulmonary Activity using A Compact Quadrature Doppler Radar Sensor. *IEEE Transactions on Biomedical Engineering*, 2013.
- [25] BITALINO Inc. Bitalino Revolution Plugged Kit BT. <https://bitalino.com/en/plugged-kit-bt>, 2020. [Online; accessed 23-August-2020].
- [26] TI Inc. TI Product. <https://www.ti.com/product/IWR1843>, 2020. [Online; accessed 23-August-2020].
- [27] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu. Smart Home Based on WiFi Sensing: A Survey. *IEEE Access*, 2018.
- [28] S. Jo, J. Kim, and D. Kim. Heart Rate Change While Drowsy Driving. *Journal of Korean medical science*, 2019.
- [29] F. Khan and S. Cho. A Detailed Algorithm for Vital Sign Monitoring of A Stationary/Non-stationary Human Through IR-UWB Radar. *Sensors*, 2017.
- [30] Y. Koo, L. Ren, Y. Wang, and A. Fathy. UWB Micro Doppler Radar for Human Gait Analysis, Tracking More Than One Person, and Vital Sign Detection of Moving Persons. In *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013.
- [31] L. Ren, H. Wang, K. Naishadham, Q. Liu, and A. E. Fathy. Non-invasive Detection of Cardiac and Respiratory Rates From Stepped Frequency Continuous Wave Radar Measurements using The State Space Method. In *IEEE MTT-S International Microwave Symposium*, 2015.
- [32] K. Lee, D. Han, and H. Ko. Video Analytic Based Health Monitoring for Driver in Moving Vehicle by Extracting Effective Heart Rate Inducing Features. *Journal of Advanced Transportation*, 2018.
- [33] C. Li and J. Lin. Random Body Movement Cancellation in Doppler Radar Vital Sign Detection. *IEEE Transactions on Microwave Theory and Techniques*, 2008.
- [34] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang. Monitoring Vital Signs and Postures During Sleep using WiFi Signals. *IEEE Internet of Things Journal*, 2018.
- [35] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng. Tracking Vital Signs During Sleep Leveraging Off-the-shelf WiFi. In *ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015.

- [36] K. J. R. Liu and B. Wang. *Wireless AI: Wireless Sensing, Positioning, IoT, and Communications*. Cambridge University Press, 2019.
- [37] Q. Lv, L. Chen, K. An, J. Wang, H. Li, D. Ye, J. Huangfu, C. Li, and L. Ran. Doppler Vital Signs Detection in the Presence of Large-Scale Random Body Movements. *IEEE Transactions on Microwave Theory and Techniques*, 2018.
- [38] N. Magdalena, T. Marks, H. Mansour, and A. Veeraraghavan. SparsePPG: Towards Driver Monitoring using Camera-based Vital Signs Estimation in Near-infrared. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018.
- [39] M. Marcus and B. Pattan. Millimeter Wave Propagation: Spectrum Management Implications. *IEEE Microwave Magazine*, 2005.
- [40] C. S. Marino and P. M. Chau. High-resolution DOA Estimation from Synthetic Aperture Beamforming. In *2005 IEEE Antennas and Propagation Society International Symposium*, 2005.
- [41] D. McDuff, S. Gontarek, and R. Picard. Remote Measurement of Cognitive Stress via Heart Rate Variability. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.
- [42] MDOT SHA Pavement Condition. <https://data-maryland.opendata.arcgis.com/datasets/mdot-sha-pavement-condition-1>.
- [43] A. Melchor Rodríguez and J. Ramos Castro. Pulse Rate Variability Analysis by Video using Face Detection and Tracking Algorithms. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
- [44] M. Mercuri, Y. Liu, I. Lorato, T. Torfs, F. Wieringa, A. Bourdoux, and C. Van Hoof. A Direct Phase-Tracking Doppler Radar using Wavelet Independent Component Analysis for Non-Contact Respiratory and Heart Rate Monitoring. *IEEE Transactions on Biomedical Circuits and Systems*, 2018.
- [45] M. Mercuri, I. R. Lorato, Y. Liu, F. Wieringa, C. Van Hoof, and T. Torfs. Vital-sign Monitoring and Spatial Tracking of Multiple People using A Contactless Radar-based Sensor. *Nature Electronics*, 2019.
- [46] MIMO Radar. <https://www.ti.com.cn/cn/lit/an/swra554a/swra554a.pdf?ts=1598454063259>.
- [47] E. Mogi and T. Ohtsuki. Heartbeat Detection with Doppler Radar Based on Spectrogram. In *IEEE International Conference on Communications (ICC)*, 2017.

- [48] I. Mostafanezhad, E. Yavari, O. Boric-Lubecke, V. M. Lubecke, and D. P. Mandic. Cancellation of Unwanted Doppler Radar Sensor Motion using Empirical Mode Decomposition. *IEEE Sensors Journal*, 2013.
- [49] N. T. Phuong Nguyen, P.-Y. Lyu, M. H. Lin, C.-C. Chang, and S.-F. Chang. A Short-Time Autocorrelation Method for Noncontact Detection of Heart Rate Variability using CW Doppler Radar. In *2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, 2019.
- [50] T. Ohtsuki and E. Mogi. Heartbeat Detection with Doppler Radar Based on Estimation of Average R-R Interval using Viterbi Algorithm. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016.
- [51] A. Pai, A. Veeraraghavan, and A. Sabharwal. HRVCam: Robust Camera-based Measurement of Heart Rate Variability. *Journal of Biomedical Optics*, 2021.
- [52] V. L. Petrovic, M. M. Jankovic, A. V. Lupsic, V. R. Mihajlovic, and J. S. Popovic-Bozovic. High-accuracy Real-time Monitoring of Heart Rate Variability using 24 GHz Continuous-wave Doppler Radar. *IEEE Access*, 2019.
- [53] Polar H10 heart rate monitors. https://www.polar.com/us-en/products/accessories/h10_heart_rate_sensor.
- [54] K. Qian, C. Wu, F. Xiao, Y. Zheng, Y. Zhang, Z. Yang, and Y. Liu. Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices. In *IEEE Conference on Computer Communications (INFOCOM)*, 2018.
- [55] Qualcomm 802.11ad 60GHz WiFi. <https://www.qualcomm.com/products/features/80211ad>.
- [56] V. R. Radzicki, D. Boutte, P. V. Taylor, and H. Lee. Standoff CW Radar for Through-the-wall Detection of Human Heartbeat Signatures. In *IEEE Radar Conference (RadarConf)*, 2016.
- [57] G. Ramachandran and M. Singh. Three-dimensional Reconstruction of Cardiac Displacement Patterns on The Chest Wall During The P, QRS and T-segments of The ECG by Laser Speckle Interferometry. *Medical and Biological Engineering and Computing*, 1989.
- [58] L. Ren, Y. S. Koo, Y. Wang, and A. Fathy. Noncontact Heartbeat Detection using UWB Impulse Doppler Radar. In *2015 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireless)*, 2015.
- [59] L. Ren, H. Wang, K. Naishadham, O. Kilic, and A. E. Fathy. Phase-Based Methods for Heart Rate Detection using UWB Impulse Doppler Radar. *IEEE Transactions on Microwave Theory and Techniques*, 2016.

- [60] M. A. Richards. *Fundamentals of Radar Signal Processing, Second Edition*. McGraw-Hill Education, 2014.
- [61] Road Traffic Injuries. https://www.who.int/health-topics/road-safety#tab=tab_1.
- [62] A. Melchor Rodríguez and J. Ramos-Castro. Video Pulse Rate Variability Analysis in Stationary and Motion Conditions. *Biomedical engineering online*, 2018.
- [63] T. Sakamoto, R. Imasaka, H. Taki, T. Sato, M. Yoshioka, K. Inoue, T. Fukuda, and H. Sakai. Feature-based Correlation and Topological Similarity for Interbeat Interval Estimation using Ultrawideband Radar. *IEEE Transactions on Biomedical Engineering*, 2015.
- [64] F. Shaffer and J. P. Ginsberg. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in public health*, 2017.
- [65] A. Sharafi, M. Baboli, M. Eshghi, and A. Ahmadian. Respiration-rate Estimation of A Moving Target using Impulse-based Ultra Wideband Radars. *Australasian physical & engineering sciences in medicine*, 2012.
- [66] H. Shen, C. Xu, Y. Yang, L. Sun, Z. Cai, L. Bai, E. Clancy, and X. Huang. Respiration and Heartbeat Rates Measurement Based on Autocorrelation using IR-UWB Radar,. *IEEE Transactions on Circuits and Systems II: Express Briefs*.
- [67] J. Solaz, J. Laparra-Hernández, D. Bande, N. Rodríguez, S. Veleff, J. Gerpe, and E. Medina. Drowsiness Detection Based on The Analysis of Breathing Rate Obtained From Real-time Image Recognition. *Transportation research procedia*, 2016.
- [68] S. Suzuki, T. Matsui, H. Imuta, M. Uenoyama, H. Yura, M. Ishihara, and M. Kawakami. A Novel Autonomic Activation Measurement Method for Stress Monitoring: Non-contact Measurement of Heart Rate Variability using A Compact Microwave Radar. *Medical & biological engineering & computing*, 2008.
- [69] R. L. Rivest T. H. Cormen, C. E. Leiserson. *Introduction to Algorithms*. MIT press, 2009.
- [70] M. Tang, F. Wang, and T. Horng. Single Self-Injection-Locked Radar With Two Antennas for Monitoring Vital Signs With Large Body Movement Cancellation. *IEEE Transactions on Microwave Theory and Techniques*, 2017.
- [71] C. Tomasi and R. Manduchi. Bilateral Filtering for Gray and Color Images. In *International Conference on Computer Vision*, 1998.
- [72] N. u. Rehman and H. Aftab. Multivariate Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 2019.

- [73] B. Wang, Q. Xu, C. Chen, F. Zhang, and K. J. R. Liu. The Promise of Radio Analytics: A Future Paradigm of Wireless Positioning, Tracking, and Sensing. *IEEE Signal Processing Magazine*, 2018.
- [74] F. Wang, X. Zeng, C. Wu, B. Wang, and K. J. R. Liu. mmHRV: Contactless Heart Rate Variability Monitoring using Millimeter-Wave Radio. *IEEE Internet of Things Journal*, 2021.
- [75] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu. ViMo: Multi-person Vital Sign Monitoring using Commodity Millimeter Wave Radio. *IEEE Internet of Things Journal*, 2020.
- [76] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. Ray Liu. Respiration Tracking for People Counting and Recognition. *IEEE Internet of Things Journal*, 2020.
- [77] J. Wang, X. Wang, Z. Zhu, J. Huangfu, C. Li, and L. Ran. 1-D Microwave Imaging of Human Cardiac Motion: An Ab-Initio Investigation. *IEEE Transactions on Microwave Theory and Techniques*, 2013.
- [78] W. Wang, A. Liu, and M. Shahzad. Gait Recognition using Wifi Signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16*, 2016.
- [79] X. Wang, C. Yang, and S. Mao. TensorBeat: Tensor Decomposition for Monitoring Multi-Person Breathing Beats with Commodity WiFi. *ACM Transactions on Intelligent Systems and Technology*, 2017.
- [80] C. Will, K. Shi, R. Weigel, and A. Koelpin. Advanced Template Matching Algorithm for Instantaneous Heartbeat Detection using Continuous Wave Radar Systems. In *2017 First IEEE MTT-S International Microwave Bio Conference (IMBIOC)*, 2017.
- [81] World Vehicles In Use. https://www.oica.net/wp-content/uploads/Total_in-use-All-Vehicles.pdf.
- [82] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao. Non-Invasive Detection of Moving and Stationary Human with WiFi. *IEEE Journal on Selected Areas in Communications*, 2015.
- [83] Q. Xu, Y. Chen, B. Wang, and K. J. R. Liu. Radio Shot: Through-the-wall Human Recognition. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016.
- [84] Q. Xu, B. Wang, F. Zhang, D. S. Regani, F. Wang, and K. J. R. Liu. Wireless AI in Smart Car: How Smart a Car Can Be? *IEEE Access*, 2020.
- [85] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra. Monitoring Vital Signs using Millimeter Wave. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2016.

- [86] Z. Yang, H. Shi, S. Zhao, and X. Huang. Vital Sign Detection During Large-scale and Fast Body Movements Based on An Adaptive Noise Cancellation Algorithm using A Single Doppler Radar Sensor. *Sensors*, 2020.
- [87] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang. FullBreathe: Full Human Respiration Detection Exploiting Complementarity of CSI Phase and Amplitude of WiFi Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018.
- [88] D. Zhang, Y. Hu, Y. Chen, and B. Zeng. BreathTrack: Tracking Indoor Human Breath Status via Commodity WiFi. *IEEE Internet of Things Journal*, 2019.
- [89] F. Zhang, C. Chen, B. Wang, and K. J. R. Liu. WiSpeed: A Statistical Electromagnetic Approach for Device-Free Indoor Speed Estimation. *IEEE Internet of Things Journal*, 2018.
- [90] F. Zhang, C. Wu, B. Wang, H. Lai, Y. Han, and K. J. Ray Liu. WiDetect: Robust Motion Detection with a Statistical Electromagnetic Model. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2019.
- [91] F. Zhang, C. Wu, B. Wang, M. Wu, D. Bugos, H. Zhang, and K. J. R. Liu. SMARS: Sleep Monitoring via Ambient Radio Signals. *IEEE Transactions on Mobile Computing*, 2019.
- [92] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere. WiFi-ID: Human Identification using WiFi Signal. In *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2016.
- [93] M. Zhao, F. Adib, and D. Katabi. Emotion Recognition using Wireless Signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016.
- [94] T. Zheng, Z. Chen, C. Cai, J. Luo, and X. Zhang. V2iFi: in-Vehicle Vital Sign Monitoring via Compact RF Sensing. 2020.
- [95] Q. Zhu, M. Chen, C. Wong, and M. Wu. Adaptive Multi-Trace Carving Based on Dynamic Programming. In *Asilomar Conference on Signals, Systems, and Computers*, 2018.
- [96] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos. Wifi-based Human Identification via Convex Tensor Shapelet Learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.